Selected methodological characteristics of meta-analyses related to educational achievement are reviewed in an exploration of the practice of meta-analysis and the characteristics of meta-analyses related to educational achievement, as well as possible relationships among background, methodological and substantive characteristics, and effect sizes. A literature search identified 1,197 documents, of which 694 were retrieved as pertinent. Using only meta-analyses published after 1984, 103 published meta-analyses were selected as having met study criteria. The most frequent type of meta-analysis was that of treatment effectiveness. Hypothesis and theory testing did not appear as frequently as descriptive research. Many primary research articles did not include sample size, precluding the computation of effect size. Many details of the search procedures in meta-analyses were not included, and fewer than 40% of the authors reported some kind of homogeneity of effect size testing. Overall, results suggest that researchers are not exploiting the full capabilities of meta-analytic techniques. Appendix A lists meta-analyses included in the study, and Appendix B lists those specifically excluded. (Contains 6 tables, 10 figures, and 38 references.) (SLD)
A META-META-ANALYSIS: METHODOLOGICAL ASPECTS OF META-AANALYSES IN EDUCATIONAL ACHIEVEMENT

by

Theresa Ann Sipe
Georgia State University

and

William L. Curlette
Georgia State University

Paper presented at AERA Annual Meeting Program
April 8-12, 1996
New York City, NY

email: tsipe@gsu.edu
ACKNOWLEDGMENTS

"The significant problems we face cannot be resolved at the same level of thinking we were at when we created them"

Albert Einstein

We would like to acknowledge the coders, Eve Ackerman, Fulya Bergman, Karen Popoff, and Gary Martin who performed an integral part of this research.

Many thanks to Herbert Walberg, Bill Stallings, and Kathi Cannella who added their expertise to this project.
A Meta-Meta-Analysis: Methodological Aspects of Meta-Analyses in Educational Achievement

Glass first coined the term meta-analysis in 1976 to imply "the analysis of analyses" (Glass, 1978, p. 352). Since 1976, meta-analyses have proliferated and hundreds now exist in the educational, psychological, and medical literatures. Although not widely discussed as a methodology, a handful of reviewers have conducted research summarizing and/or synthesizing meta-analytic studies. Reviews of this nature were located in the areas of education (Abrami, Cohen, & d'Apollonia, 1988; Anderson, 1983; Cooper, Dorr, & Bettencourt, 1995; Fraser, Walberg, Welch, & Hattie, 1987; Hattie, 1991; Kulik & Kulik, 1987; Kulik & Kulik 1989; Rosenthal, 1991; Wang, Haertel, & Walberg, 1993), special education (Kavale & Dobbins, 1993; Swanson et al., 1993), psychology (Cornwell, 1987; Cornwell, 1988; Lipsey & Wilson, 1993), industrial and organizational psychology (Hunter & Hirsh, 1987), communication (Canary & Hause, 1993), and medicine (Emerson, Burdick, Hoaglin, Mosteller, & Chalmers, 1990; Sacks, Berrier, Reitman, Ancona-Berk, & Chalmers, 1987). The purposes of these reviews vary from summarizing substantive or methodological characteristics to assessing a theoretical model.

Objectives

Although a few of the reviewers listed above have examined methodological, substantive, or outcome characteristics of meta-analyses, the reviews are either dated, limited in scope, contain a small sample size, or are conducted in areas other than
education. As a result, the state of meta-analysis in education has not been fully documented. The purpose of this research project was threefold. The first goal involved testing a model of learning [viz., the model of educational productivity developed by Walberg (1984)] on those meta-analyses related to achievement. The second involved describing background, methodological, and substantive characteristics of meta-analyses in education which are related to achievement. The third goal involved exploration of possible relationships among background, methodological, and substantive characteristics and effect sizes. In this paper, selected methodological characteristics of meta-analyses related to educational achievement will be presented. In addition, the curriculum interventions with the 5 highest and 5 lowest effect sizes will be identified.

Perspective

Glass, McGaw and Smith defined meta-analysis as the "attitude of data analysis applied to quantitative summaries of individual experiments" (1981, p. 21). Hedges and Olkin (1985) referred to meta-analysis as the "analysis of the results of statistical analyses (p. 13). Meta-analysis is an "orientation" of combining research studies that uses a variety of techniques of measurement and data analysis (Wachter & Straf, 1990). Meta-analytic techniques have the ability to go beyond simple vote-counting across studies and evaluate the conditions under which effects occur as well as explore the mediating processes that may underlie those effects (Cook et al., 1992).
As noted earlier, a handful of researchers have taken literature integration one step further by summarizing or synthesizing meta-analyses. Eighteen studies which used this methodological approach were found in the literature. However, this approach to literature summarization has had very little discussion as a methodology by the either scholars who have conducted this type of literature review or other experts in the area of literature summarization. The methodology has remained implicit and secondary to the authors' primary focus of obtaining particular information from meta-analytic studies. Consequently, this review of the literature will only contain applications of this methodology but no in-depth discussion of the methodology.

Review of Studies Using Meta-Analyses as the Unit of Analysis

The 18 studies ranged in size, purpose, and scope. Three of the 18 documents shared a data set and were thus not unique (viz., Kulik & Kulik, 1987 with Kulik and Kulik, 1989; Cornwell, 1987 with Cornwell, 1988; and Fraser et al., 1987 with Hattie, 1991). Twelve of the studies can be categorized as summarizations of certain variables across meta-analyses while six of the studies went beyond summarization and synthesized findings within a theoretical framework or conducted hypothesis testing (see Table 1). The latter is defined in this study as meta-synthesis. As seen in Table 1, seven studies had a sample size less than 10, five studies had a sample size greater than 10 but less than 100, four studies had a sample size greater than 100, and two studies did not report sample size. Sacks et al. (1987) collected methodological data only. Canary and Hause (1993) and Kavale and
Table 1

Synopsis of Studies Using Meta-Analyses as the Unit of Analysis

<table>
<thead>
<tr>
<th>Type of Statistics</th>
<th>n\textsuperscript{a}</th>
<th>Descriptive</th>
<th>Other\textsuperscript{b}</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson (1983)</td>
<td>7</td>
<td>X</td>
<td></td>
<td>Summary</td>
</tr>
<tr>
<td>Cornwell (1987, 1988)</td>
<td>81</td>
<td>X</td>
<td></td>
<td>Summary</td>
</tr>
<tr>
<td>Hunter &amp; Hirsh (1987)</td>
<td>?</td>
<td>X</td>
<td></td>
<td>Summary</td>
</tr>
<tr>
<td>Sacks et al. (1987)</td>
<td>86</td>
<td>X</td>
<td></td>
<td>Summary</td>
</tr>
<tr>
<td>Abrami et al. (1988)</td>
<td>6</td>
<td>X</td>
<td>X</td>
<td>Summary</td>
</tr>
<tr>
<td>Kulik &amp; Kulik (1987)</td>
<td>4</td>
<td>X</td>
<td>X</td>
<td>Summary</td>
</tr>
<tr>
<td>Kulik &amp; Kulik (1989)</td>
<td>?</td>
<td>X</td>
<td></td>
<td>Summary</td>
</tr>
<tr>
<td>Rosenthal (1991b)</td>
<td>8</td>
<td>X</td>
<td></td>
<td>Summary</td>
</tr>
<tr>
<td>Canary &amp; Hause (1993)</td>
<td>15</td>
<td>X</td>
<td></td>
<td>Summary</td>
</tr>
<tr>
<td>Kavale &amp; Dobbins (1993)</td>
<td>6</td>
<td>X</td>
<td></td>
<td>Summary</td>
</tr>
<tr>
<td>Swanson et al. (1993)</td>
<td>3</td>
<td>X</td>
<td></td>
<td>Summary</td>
</tr>
<tr>
<td>Fraser et al. (1987)</td>
<td>134</td>
<td>X</td>
<td></td>
<td>Synthesis</td>
</tr>
<tr>
<td>Emerson et al. (1990)</td>
<td>7</td>
<td>X</td>
<td>X</td>
<td>Synthesis</td>
</tr>
<tr>
<td>Hattie (1991)</td>
<td>134</td>
<td>X</td>
<td></td>
<td>Synthesis</td>
</tr>
<tr>
<td>Lipsey &amp; Wilson (1993)</td>
<td>302</td>
<td>X</td>
<td></td>
<td>Synthesis</td>
</tr>
<tr>
<td>Cooper et al. (1995)</td>
<td>302</td>
<td>X</td>
<td></td>
<td>Synthesis</td>
</tr>
<tr>
<td>Wang et al. (1993)</td>
<td>91</td>
<td>X</td>
<td>X</td>
<td>Synthesis</td>
</tr>
</tbody>
</table>

\textsuperscript{a} n of meta-analytic studies

\textsuperscript{b} Other types of statistics included correlation, regression, other linear models, t test, and chi square.
Dobbins (1993) reported substantive data only. Cornwell (1988), Abrami et al. (1988), and Emerson et al. (1990) collected methodological and outcome data but no substantive data. Anderson (1983), Hunter and Hirsh (1987), Kulik and Kulik (1987), Kulik and Kulik (1989), Rosenthal (1991b), and Swanson et al. (1993) examined methodological and substantive variables. Fraser et al. (1987) integrated many meta-analyses to assess the model of school learning. Hattie (1991) extended the work done in Fraser et al. (1987) by converting the correlational effect sizes to mean effect sizes and introducing the notion of a universal continuum to measure the effects of schooling. Lipsey and Wilson (1993) grouped meta-analyses by topic area but did not integrate them into any theoretical model. Cooper et al. (1995) did a secondary analysis on the data reported by Lipsey and Wilson (1993). Wang et al. (1993) compiled content analyses, expert ratings, and meta-analyses to assess a theoretical framework of school learning. Although Fraser et al. (1987), Lipsey and Wilson (1993), and Wang et al. (1993) were the most comprehensive studies of this nature, only Wang et al. reported statistical analyses beyond descriptive statistics. This research project differs from the above reviewed studies in recency and scope. Data were collected for a comprehensive set of background, methodological, and substantive variables from meta-analyses published between 1984 and 1993.

Methods

The methods used in this project were similar to the stages of research synthesis suggested by Cooper and Hedges (1994) which
encompassed problem formulation, data collection, data
evaluation, analysis and interpretation. The design of this
research is a "meta-synthesis," a new term defined by the authors
as the summarization and synthesis of meta-analytic studies. Data
collection included the literature search and retrieval process
described below. Data evaluation consisted of coding the meta-
analyses. A five part coding form and manual were developed for
this project. The coding form was pilot tested and revisions were
made. A team of two coders coded a total of 46 methodological
variables. Over the nine month period it took to complete this
portion of the project, the coders met weekly to discuss and
resolve discrepancies for each document. Unresolved discrepancies
were resolved by the first author. Interrater reliability was
established using the first 5 and last 5 meta-analyses. Percent
agreement was 90.85% and 91.62% respectively with an average of
91.2% across all 10 studies. Content validity was achieved
through a review process involving several experts in meta-
analysis as well as experts in education. Several revisions were
made based on recommendations from the experts. Establishing
criterion validity was beyond the scope of this research project.

Data Source

Using the suggestions by Cook et al. (1992) and Cooper
(1989), the literature search procedures were conducted in
several steps that included obtaining citations from the
following sources: computerized database searching of ERIC and
PsycLIT; ancestry; invisible college; personal readings; and hand
searching of the Review of Educational Research. The keywords:
("meta-analytic" or "meta analytic" or "meta-analysis or "meta analysis" or "quantitative synthesis" or " Best Evidence Synthesis") and ("education" or "coaching" or "training" or "teaching" or "achievement") and "language = English," identified a total of 1197 citations (see Table 2).

Table 2

Numbers of Citations Resulting From All Searching Strategies

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of Citations Identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERIC</td>
<td>752</td>
</tr>
<tr>
<td>PsychLIT</td>
<td>335</td>
</tr>
<tr>
<td>Ancestry</td>
<td>42</td>
</tr>
<tr>
<td>Personal Readings</td>
<td>36</td>
</tr>
<tr>
<td>Invisible College</td>
<td>26</td>
</tr>
<tr>
<td>Hand Search of Journal</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>1197</td>
</tr>
</tbody>
</table>

Once citations were identified, titles and abstracts were read to determine if retrieval was necessary. A total of 694 documents were retrieved.

Criteria for Inclusion of Meta-Analytic Studies

The criteria for inclusion of meta-analyses in the study were:
1. Published journal articles for the years 1984-1993; [Several authors recommend inclusion of unpublished studies in meta-analysis since unpublished studies are more likely to have nonsignificant results, and consequently, lower effect sizes (Cook et al., 1992; Glass et al., 1981; Rosenthal, 1991a). However, Cooper et al. (1995) found published and unpublished meta-analyses differed by no more than 0.04 standard deviations in effect size.]

2. Published research reports of meta-analyses when the corresponding meta-analysis was not included in the database;

3. Meta-analyses with at least one outcome measure of achievement; and

4. At least one reported effect size or statistic which could be converted into an effect size.

Criteria for Exclusion of Meta-Analytic Studies

The criteria for exclusion of studies from the meta-synthesis were:

1. Outcomes with higher education programs in certain fields of study (e.g., medical, nursing, dental) since these represent areas in a specialized knowledge base;

2. Outcomes with preschoolers because achievement measures are different at this level;

3. Aptitude outcomes since these measure the ability to perform rather than achievement; and

4. Interventions with high risk infants.

Four hundred twenty-seven meta-analyses were among the citations retrieved. One hundred seventeen meta-analyses were
published prior to 1984 and thus did not meet criteria for inclusion in the study. One hundred ninety-one published meta-analyses did not have outcome measures related to achievement and were also not included in the study. The remaining 119 published meta-analyses were related to achievement and identified for possible inclusion in the study. Of these, 16 meta-analyses were excluded based on the exclusion criteria listed above (see Appendix B). Therefore, 103 published meta-analyses were included in the study (see Table 3 and Appendix A).

Table 3
Source of Meta-Analyses Included in the Study

<table>
<thead>
<tr>
<th>Source of Meta-Analyses</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer search</td>
<td>100</td>
</tr>
<tr>
<td>Ancestry</td>
<td>2</td>
</tr>
<tr>
<td>Invisible college</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>103</td>
</tr>
</tbody>
</table>

Data Management

All data were double-entered and validated using Epi Info Version 6.02, which is a word processing, database, and statistics system for Epidemiology on Microcomputers (Dean, Dean, Burton, & Dicker, 1990). Statistical analysis was done through Epi Info Version 6.02, SPSSx (version 4.1) for the mainframe
Graphs were produced using SPSS® for Windows™ (SPSS Inc., 1994) and Quatro® Pro for Windows Version 5 (Borland International, Inc., 1993).

Research Questions

1. What are the methodological characteristics included in educational meta-analyses related to achievement?

2. Are there any relationships between the methodological characteristics and mean effect size from each meta-analysis related to achievement?

3. What are the curriculum interventions with the 5 highest and lowest effect sizes?

Data Analysis

The standardized difference between group means was the effect size metric used in all statistical computations. Correlational effect sizes were converted from an r to a Cohen's d which is a standardized difference between group means (Rosenthal, 1994, p. 239)

\[ d = \frac{2r}{1/2(1-r^2)}. \] (1)

Descriptive statistics were computed to answer the research questions. A sensitivity analysis was conducted to explore possible relationships among various variables and effect sizes which could indicate publication bias (Begg, 1994; Greenhouse & Iyengar, 1994). Publication bias occurs when studies with statistical significance are published and studies with no significant results are not published (Greenhouse & Iyengar,
Data were analyzed using funnel plots, the correlation coefficients Kendall's tau and Spearman's rho (Begg, 1994; Greenhouse & Iyengar, 1994), as well as descriptive statistics. Light and Pillemer (1984) first introduced funnel plots to detect publication bias. Funnel plots are scatterplots of sample size versus mean effect size. Larger studies generally have less variability among the effect sizes than smaller studies (Greenhouse & Iyengar, 1994). Since there are usually more smaller studies, the plot should look like a funnel (Greenhouse & Iyengar, 1994). Begg (1994) suggested using Kendall's tau correlation test which has low power but involves no modeling assumptions. Alternately, Begg (1994) suggested conducting Spearman's rho which is based on the assumption that "that the effect sizes are statistically independent and identically distributed, under the null hypothesis of no bias" (p. 402). Both correlation coefficients were computed for this project. Before computing the Spearman's rho, the mean effect sizes ($d$) were transformed first into correlations (Rosenthal, 1994, p. 239)

$$r = \frac{1}{2}(d^2 / (d^2 + 4)). \tag{2}$$

Next, the correlations were transformed into $Z_r$ (Rosenthal, 1994, p. 240)

$$Z_r = \frac{1}{2} \log_e \frac{(1 + r)}{(1 - r)} \tag{3}$$

Results

Results for selected methodological variables as well the curriculum interventions with the 5 highest and 5 lowest effect sizes are presented below.
Methodological Variables

Purpose of meta-analysis. The most common purpose for the meta-analyses was that of investigating treatment (Rx) effectiveness (see Figure 1). Variable covariation refers to meta-analyses in which relationships among variables were explored. The category of other included replication of another meta-analysis, testing construct validity of cognitive preferences, testing construct validity of computer based instruction, conducting a review of process product research, or theory building. One meta-analysis (1%) involved test validity, which refers to a type of meta-analysis that explores the correlation between a test or measure and a criterion variable (Durlak & Lipsey 1991).

Eighty-two percent of the authors reported a singular purpose which was most frequently treatment effectiveness. Eighteen percent reported a dual purpose which involved treatment effectiveness and another purpose. The theories or explanatory mechanisms tested in nine of the meta-analyses are listed in Table 4.

Research questions. The most frequent type of research question or questions was descriptive in nature (see Figure 2). The type of research question for one meta-analysis (1%) was unknown. The one meta-analysis categorized as other reported the research question as theory building. Eighty-one percent of the authors reported a singular type of research question or hypothesis.
Figure 1. Purposes of the meta-analyses. Meta-analyses may contain more than one purpose.

Meta-analytic procedures. The Glass meta-analytic procedures were the most commonly used procedures reported followed by the Hedges meta-analytic procedures (see Figure 3). Five meta-analyses (4.9%) contained Rosenthal meta-analytic procedures. About half of the authors reported using a combination of Glass, Hedges, and some other procedure.

Search procedures. Eighty-four percent of the meta-analyses contained some details concerning the literature searching procedures used in identifying primary studies for the meta-
Table 4

List of Theories or Explanatory Mechanisms Tested in the Meta-Analyses

<table>
<thead>
<tr>
<th>Theory or Explanatory Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deprivation theory</td>
</tr>
<tr>
<td>Encoding hypothesis</td>
</tr>
<tr>
<td>Encoding specificity hypothesis</td>
</tr>
<tr>
<td>External storage hypothesis</td>
</tr>
<tr>
<td>Information processing</td>
</tr>
<tr>
<td>Mastery learning theory</td>
</tr>
<tr>
<td>Modality model</td>
</tr>
<tr>
<td>Motivation theory of instructional learning</td>
</tr>
<tr>
<td>Motivational theory</td>
</tr>
<tr>
<td>Only child uniqueness</td>
</tr>
<tr>
<td>Parent child relationship</td>
</tr>
<tr>
<td>Self-efficacy theory</td>
</tr>
<tr>
<td>Test anxiety</td>
</tr>
<tr>
<td>Test expectancy effect</td>
</tr>
</tbody>
</table>

Note. None of the meta-analyses contained more than one theory or explanatory mechanism.

analysis while 15.5% contained no details. The most common method of literature searching was ancestry (68%) in which new citations were identified from reference lists of obtained documents.
Sixty-six of the authors reported utilizing computers in the search procedure. Twenty percent reported manual search of paper indexes. Eighteen percent of the authors reported searching through the contents of specified volumes of certain journal(s). Nine meta-analyses (8.7%) were categorized as other for methods such as hand searching textbooks, author identified sources, author’s own studies, and data from a research center or testing agency. Seven percent of the authors reported the practice of obtaining documents from experts in the field or through conferences and workshops which has been referred to as the invisible college by Cooper (1989).

![Bar Chart](chart.png)

**Figure 2.** Types of research questions or hypotheses. Meta-analyses may contain more than one type of question.
Figure 3. Types of meta-analytic procedures reported. Meta-analyses may contain more than one type of procedure.

Details of search procedures. Eighty-six percent of the authors reported the inclusion criteria used in identifying the primary studies for the meta-analyses. Sixty-eight percent reported the citations for those primary studies included in the meta-analyses. Twenty-nine percent of the authors reported the beginning years of the literature search while 26% reported the ending years of the literature search which spanned from 4 to 36 years (mean = 15.56 years). Twenty-seven percent of the authors provided the keywords used in the computerized or manual search.
procedures. Five percent of the authors reported citations for primary studies excluded from the meta-analyses.

Ninety-three percent of the authors reported the number of primary studies included in the meta-analysis and 60% reported the number effect size measurements that were calculated. However, 41% of the authors reported the number of studies resulting from the search procedures while 39% of the authors reported the number of studies excluded from the meta-analysis. Sixty-eight percent of the meta-analyses contained unpublished primary studies.

Variables coded. Forty-three percent of the authors reported the number of variables coded in the meta-analyses which ranged 5 to 183. Eighty-eight percent of the authors reported coding substantive variables while 82% of the authors reported coding methodological variables in the meta-analysis. Twenty-nine percent of the authors listed all of the substantive and methodological variables which were coded.

Bias. Forty-six percent of the authors reported coding internal validity variables while 22% reported assessing quality of the primary research. Twenty percent of the authors reported that there were two coders per primary study. One meta-analysis contained 3 coders per primary study. The remainder either reported one coder per primary study or did not report any information about the number of coders.

Twenty-two percent of the authors reported an interrater reliability coefficient. Eighteen authors reported percent agreement, one author reported an alpha coefficient, one author
reported kappa, and one author reported Scott's \( \pi \) coefficient. One author reported that the coders were blinded to the primary research documents through a photocopy process.

**Outliers.** Twenty-six percent of the authors reported the presence of outliers in the meta-analysis. The methods used to handle the outliers are shown in Figure 4. Outliers included and outliers excluded refer to computing analyses with and without outliers. Outliers examined refers to the examination of outliers qualitatively to discern more information about them. Outliers used refers to the use of outlier effect sizes to make

![Outlier Disposition](image)

**Figure 4.** Report of the various ways outliers were handled in the meta-analyses. Note that each bar represents the percentage of the meta-analyses with outliers reported.
suggestions for further research. Outliers clustered refers to clustering outliers into separate categories for analyses.

Discussion section. The topics addressed in the discussion section can be seen in Figure 5. Research Directions refers to directions identified from the findings of the meta-analyses.

![Bar Chart: Discussion Topics](#)

**Figure 5.** Topics addressed in the discussion section. The meta-analyses may contain more than one topic in the discussion section.

**Issues related to Type I and Type II errors.** Confidence intervals around the mean effect size were reported in 22% of the meta-analyses while Fail Safe N was reported in 9% of the meta-analyses. Fail Safe N is the number of studies needed to reverse
the conclusion of significant results (Cooper, 1984). One author (1%) reported conducting a power analysis.

**Heterogeneity of effect sizes.** Fourteen percent of the authors computed a Hedges $h$ Statistic which was the precursor to the Hedges and Olkin $Q$ statistic. Thirteen authors computed the $Q$ Statistic which is used to determine whether studies share a common effect size before pooling the effect sizes for statistical analyses (Hedges & Olkin, 1985). One author reported the Hunter & Schmidt 75% Rule which refers to a rule of thumb proposed by Hunter, Schmidt, and Jackson (1990). In any data set correctable artifacts account for 75% of the variance in study correlations and uncontrolled artifacts account for the remaining 25%. Consequently, if the sampling error is 75% or more of the total variance, one can assume that the residual does not represent meaningful variation among effect sizes (Hunter et al., 1990).

**Statistical analyses.** Types of statistical analyses conducted in the meta-analyses can be seen in Figure 6. ANOVA and $t$ test were the two most common analyses reported. Regression refers to multiple regression as well as weighted least squares regression. Hedges ANOVA refers to Hedges (1982) chi-square analogue to analysis of variance. $Q$ statistic refers to Hedges and Olkin (1985) $Q$ statistic.

**Relationship Between Methodological Variables and Effect Size**

**Sample size.** Three funnel graphs which depict the number of effect size calculations (subanalyses), number of primary studies, and number of persons plotted against the total mean
effect size are seen in Figures 7, 8, and 9 respectively. One study was eliminated from the funnel graph depicting number of subjects versus total mean effect size because the sample size of 231223 was so large the graph was distorted.

Statistical Analyses

Figure 6. Types of statistical analyses. Meta-analyses may contain more than one type of analyses.

Number of variables coded. The number of variables coded in each meta-analysis was plotted against mean effect size. Figure 10 presents the funnel graph of these variables. Both the Kendall's tau and the Spearman's rho correlation coefficients were computed for variables of number of number of effect size calculations (subanalyses), number of primary studies, and number of persons, and number of variables plotted against the total
mean effect size. No significant relationships were found with either the Kendall's tau or the Spearman rho correlation coefficients.

![Funnel plot of the number of subanalyses versus mean effect size.](image)

**Figure 7.** Funnel plot of the number of subanalyses versus mean effect size.

**Highest and Lowest Effect Sizes**

The unweighted average of the total mean effect sizes (TMES) was .342 (SD = .293) when all the meta-analyses were included and .329 (SD = .306) and when only the unique meta-analyses were included. The curriculum interventions with the highest and lowest effect sizes are presented in Tables 5 and 6. Vocabulary
instruction had the highest effect size while ability grouping had the lowest.

Figure 8. Funnel plot of the number of primary studies versus mean effect size.

Conclusions

Methodological Variables

It is not surprising that the most frequent type of meta-analysis is that of treatment effectiveness since most research in education involves the investigation of various treatments. Although some of the goals of meta-analysis are to conduct hypothesis testing and theory testing, it appears that these types of research do not appear in meta-analyses as frequently as
descriptive research. It would be interesting to look at this variable over time.

![Funnel plot of the number of subjects versus mean effect size.](image)

**Figure 9.** Funnel plot of the number of subjects versus mean effect size.

Most authors reported a mean difference effect size while only a handful of authors reported a correlational effect size. Although there has been much controversy over the type of meta-analytic procedure used in a meta-analysis, most authors reported the Glass procedure. Many primary research articles do not include sample size and this prevents the computation of the Hedges effect size. Many of the authors that computed both the
Glass and Hedges effect size reported that the difference between the two were negligible.

![Figure 10. Funnel plot of the number of variables coded versus mean effect size.](image)

One of the advantages of meta-analysis touted by the experts is that replication with this methodology is feasible since the authors use protocols to search and gather studies. However, many details of the search procedures in the meta-analyses were not reported. This may be due to several factors such as the form in which authors elect to report studies, editorial decisions, or amount of information available in the primary studies. The implications include the possibility that replicating meta-
Table 5

List of Curriculum Interventions with Highest Mean Effect Size (MES)

<table>
<thead>
<tr>
<th>Intervention</th>
<th>n</th>
<th>MES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary instruction</td>
<td>52</td>
<td>1.147</td>
</tr>
<tr>
<td>Accelerative instruction</td>
<td>13</td>
<td>0.880</td>
</tr>
<tr>
<td>Mastery learning</td>
<td>25</td>
<td>0.821</td>
</tr>
<tr>
<td>Direct instruction</td>
<td>19</td>
<td>0.820</td>
</tr>
<tr>
<td>Notetaking</td>
<td>21</td>
<td>0.710</td>
</tr>
</tbody>
</table>

*n of primary studies

analyses may be difficult due to lack of information provided in the original meta-analysis. In addition, only limited information from the primary studies may be available to the researcher who conducts a meta-analysis.

It is not surprising that almost three-quarters of the authors addressed future research directions identified from the findings of the meta-analyses. However, less than 40% of the authors used the findings to address policy implications even though meta-analytic techniques are capable of providing information to make such implications. Less than 20% of the authors addressed the results in relation to theories which probably is related to the number of meta-analyses which test theories. The most disappointing finding is that less than 15% of
Table 6

**List of Curriculum Interventions with Lowest Mean Effect Size (MES)**

<table>
<thead>
<tr>
<th>Intervention</th>
<th>n</th>
<th>MES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability grouping</td>
<td>20</td>
<td>-.038</td>
</tr>
<tr>
<td>Frostig program</td>
<td>47</td>
<td>0.019</td>
</tr>
<tr>
<td>Matched teacher/student cognitive style</td>
<td>5</td>
<td>0.030</td>
</tr>
<tr>
<td>Factual adjunct questions</td>
<td>47</td>
<td>0.076</td>
</tr>
<tr>
<td>Intermediate Science</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curriculum Study</td>
<td>10</td>
<td>0.090</td>
</tr>
<tr>
<td>Whole language</td>
<td>34</td>
<td>0.090</td>
</tr>
</tbody>
</table>

*n of primary studies

the authors discussed the economic impact of their findings. Meta-analytic techniques can be powerful if used to their fullest potential. Encouraging authors of primary research as well as meta-analyses to consider ways of incorporating economic and policy considerations in their research is one implication from these findings.

Less than 40% of the authors reported some type of homogeneity of effect size testing. Although some authors recommend a Hedges' chi-square analog to ANOVA over the traditional ANOVA (Hedges, 1982; Hedges & Olkin, 1985), 35% of
the authors reported conducting ANOVA and only 10% reported computing a chi-square analog to ANOVA. It would appear that most authors of meta-analyses are reporting descriptive results. One implication is to encourage authors of meta-analyses to explore moderator variables.

Although some of the goals of meta-analysis are to conduct hypothesis testing and theory testing, it appears that these types of research do not appear in meta-analyses as frequently as descriptive research. In addition, the minority of meta-analysts link the meta-analysis to theory in the discussion section. These results suggest that authors are not exploiting the full capabilities of meta-analytic techniques.

While large databases of meta-analyses have been created in the past, the database from this project differs in that many methodological characteristics were coded and documented. These results provide a view of current practice in meta-analytic research for researchers who wish to utilize the meta-analytic approach. In addition, the results can be useful in teaching students and others about the methodological aspects of research synthesis.

**Highest and Lowest Effect Sizes**

The overall average effect size can provide a benchmark against which new programs can be compared. These and other findings from this meta-synthesis provide educators and researchers with an updated resource of what curriculum interventions work best in education as well as identifying those that produce poor results.
References


APPENDIX A

Meta-Analyses Included in the Study


APPENDIX B

Meta-Analyses Excluded from the Study


