

# Integrating Research into Instructional Practice: The Use and Abuse of Meta-Analysis

Robert M. Bernard  
Som Naidu

---

**Abstract:** This article sketches a broad picture of meta-analysis, a technique for quantitatively summarizing research studies. Its overall purpose is to guide researchers and practitioners working in the instructional media and technology field, in future research design and decision-making for instructional development. The article has five main sections: 1) a general introduction to the other sections describes the place of meta-analysis within the educational technology field; 2) a discussion of the reason for and the nature of integrative review as a whole, plus details of some of the objections in principle and in practice to quantitative synthesis; 3) a description of the process of conducting a meta-analysis along with a discussion of major methodological issues; 4) an example of three meta-analyses produced on one instructional treatment (i.e., mastery learning), and an explanation of different results obtained in each case; and 5) a table of 26 meta-analyses on studies of instructional variables along with guidance on how to read and interpret them.

**Résumé:** Cet article dessine un portrait de la méta-analyse qui est une technique qui résume de façon quantitative des études de recherche. La méta-analyse a pour but de conduire les chercheurs et les praticiens, qui travaillent dans le domaine de la technologie et dans les media éducatifs, à concevoir des recherches et résoudre des problèmes pratiques dans le domaine de la technologie éducative. L'article est divisé en cinq parties: 1) l'instruction générale des autres parties décrivant la position de la méta-analyse dans le domaine de la technologie éducative; 2) discussion sur la nature et l'importance d'une révision intégrale et présentation des arguments contre les synthèses quantitatives; 3) description des étapes de réalisation d'une méta-analyse; 4) un exemple de trois méta-analyses (par exemple: un apprentissage de virtuosité et une explication des différents résultats obtenus dans chaque cas; et 5) une table de 26 méta-analyses sur les variables d'instruction avec des conseils sur ses lectures et ses interprétations.

---

## INTRODUCTION

In 1983 Richard E. Clark stunned many people in the media and technology field by declaring that instructional media have no more effect on student learning and achievement than a delivery truck has on the quality of goods it transports to market. Both, he argued, are essentially neutral carriers of their respective contents. His claim extends from televised instruction on through to more recent applications of computer-based learning.

Clark's characterization of television as a neutral medium did not come as a particular shock to most, because of the results of experiments in the 1950s

---

and 1960s where no significant differences in TV treatments abounded (Saettler, 1968). But to challenge the literature of computers in education (see Clark, 1985a, 1985b) was to contradict both intuition and the prevailing research evidence. A flurry of comments and counter-comments in the literature (e.g., Petkovitch & Tennyson, 1984) and at conferences followed Clark's article for several years.

Clark's claim was based in part on an evaluation of several meta-analyses that have appeared in recent years on the effectiveness of computer-based instruction. In particular he argued that these quantitative summaries (see Table 2 at the end of this article for references) were fundamentally flawed, because a variety of experimental artifacts — among them the novelty effect associated with the treatment itself—had not been factored out of the results.

This article is about meta-analysis and its usefulness to practitioners for planning and predicting the outcomes of instructional treatments and to researchers for conceptualizing future research efforts. Meta-analysis, also referred to as quantitative synthesis, is a general set of procedures for combining the results of many individual research studies addressing a single question (Glass, 1976, 1978). The technique has grown out of a need in the social sciences to capture the essence of ever expanding research literatures and to provide definitive answers, in terms of the magnitude of effectiveness, to the bigger questions posed by theoreticians and practitioners. In addition, meta-analysis attempts to circumvent the subjectivism commonly associated with narrative forms of literature review and the limitations ascribed to the box score or vote count technique (Kavale, 1984). However, meta-analysis is not without its critics. There is disagreement among researchers on both the underlying premises of the technique as well as procedural issues relating to its implementation.

This article examines meta-analysis as a technique for reviewing literature with a particular focus on the literature of instructional techniques, methods and strategies. Some of the main issues on both sides of the "meta-analytic debate" are examined, for the purpose of judging its usefulness to educational technology and specifically its potential as a tool for designing instruction.

A case where great controversy has arisen over meta-analytic findings will then be reviewed in detail: the debate for and against the use of mastery learning. Finally, some guidance for reading and interpreting meta-analyses will be provided. An appendix to this article includes references to meta-analytic studies of instructional variables and strategies that are likely to be of interest to the educational technologist.

## METHODS OF INTEGRATIVE REVIEW

### ***Why Integrate Research Studies?***

It has long been recognized that the result of a single research study by itself is far from conclusive, even when the finding supports the hypotheses

under consideration. Therefore, it has been common practice for researchers to review the literature of all such studies, whenever enough are available. It is not uncommon, in fact, to see the same question asked and answered in reviews every couple of years, as new studies add to the weight of evidence that can be brought to bear on a particular question.

The value of integrative reviews stems from limitations that are inherent in the research process itself. Since few studies of educational phenomena and even fewer studies of instructional methods actually draw subjects at random from a population, integrative reviews of many similar studies serve to provide greater coverage of the population. The need for wider coverage is increased when one realizes that individual samples suffer from the same problems of error that is involved in testing the null hypothesis within a study. Even when a treatment effect is weak, five findings of significant differences out of one hundred studies run will be expected in the population (i.e., when  $\alpha = .05$ ), simply as a result of chance. Integrative reviews, therefore, provide a means of overcoming the effects of chance fluctuation within samples, leading to a more generalizable conclusion concerning an effect.

### ***Methods of Integrating Findings***

Light and Smith (1971) provide a typology for categorizing most reviews of research in the social sciences. The first type of review involves listing or describing factors which have produced significant differences in at least one study. The style of this type of review is primarily narrative. In the second type only studies that support a particular point of view are presented. Most of the brief reviews of literature at the beginning of research articles are of either the first or the second type. A third type involves summarizing the findings of many studies using what has come to be called the vote count or box score technique. A simple count of studies reporting positive, negative or no significant results is conducted and a verdict is reached when a plurality of votes exists. The last type, reviews in which effect sizes are aggregated across many studies, is the category in which meta-analysis resides.

The third and the fourth types of reviews are both quantitative in nature. The box score or vote count technique, however, has been criticized because it fails to take into account the effects of differential sample size on the sensitivity of the null hypothesis test. Larger samples require smaller mean differences to establish significance than smaller samples, although they are given equal credence in this technique. Box score analysis also does not take into account the magnitude of differences/relationships or the quality of the study. Meta-analysis, the subject of this article, was developed by Glass (1976, 1978) to overcome the difficulties inherent in descriptive reviews and the problems associated with using statistical indices (e.g.,  $r$ ) to reflect differential treatment effects or relationships among variables.

### ***Objections to Quantitative Synthesis***

***Objections in principle.*** Complaints about quantitative synthesis range

from the purely philosophical to the purely methodological. On the one hand, there are the arguments raised by advocates of qualitative and naturalistic approaches to enquiry (e.g., see Guba, 1979). His objections in principle to "reduction by numbers" applies doubly to quantitative synthesis, since the distillation of many studies removes the researcher one step further from the "texture" of the original setting. This, of course, is an objection that cannot be overcome by improving the theory or practice of quantitative synthesis. Rather, one must accept or reject this argument based on other criteria which have been laid out and vigorously defended by both sides.

The second major objection applies not to quantitative analysis in general, but to quantitative synthesis in particular. Eysenck (1984) argues that what is lost when narrative review is transformed to quantitative review is the exercise of scientific judgment over what is, in nearly all areas of research, a complex set of interacting variables. He states that "No simple addition of diverse and incommensurate studies can serve the purpose of drawing meaningful conclusions from heterogeneous and complex data. That requires experience, knowledge and the intangible quality we call good judgment" (p. 47).

Evidence that paints a rather different picture of actual review practice is supplied by Jackson (1980). He examined a random sampling of narrative reviews from the social sciences and found that decision rules were so often unstated in these reviews, that it was difficult to describe them, much less evaluate their quality. Meta-analysis is often touted as the antidote to the subjectivism that appears to be endemic to the process of describing research outcomes verbally.

**Objections in practice.** Hardly anyone from within the quantitative research community argues about the need to synthesize the results of large bodies of research literature. It is readily acknowledged that when the literature exceeds even a dozen studies, the ability of reviewers to capture its essence in narrative form is diminished. Quantitative synthesis, then, as a principle-for dealing with substantial literature bases is not challenged. In addition, it is generally acknowledged that there is nothing objectionable to the statistical underpinnings of meta-analysis, given that they derive from the wealth of statistical experience that has developed over many decades. It appears, then, that the objections arising from the research community derive more from the practice of meta-analyzing, than the principle of meta-analyzing.

Slavin (1984) and others have argued that one of the chief problems inherent in much of the meta-analytic literature that has appeared since its introduction by Glass, is the uncritical combining of studies that have little more in common than the underlying question -  $Y_B$  better than  $Y_C$ . It is not surprising that many practitioners have adopted this strategy and that the literature reflects this tendency, since Glass had originally suggested subjecting all available studies to meta-analysis.

In large measure, this point is at the heart of Clark's objections to the meta-analyses on computer-based learning. In this particular case, according to

Clark, the unconsidered effects of treatment artifacts produced an over-estimation of effect size for this medium of instruction, although it could have just as easily gone the other way. In any case it is generally agreed that lumping all possible forms of treatment and methodological variations into one analysis probably leads to more confusion than clarity.

It has been argued, moreover, that one of the strengths of meta-analysis, its tendency towards summary conclusions, is also one of its weaknesses (Guskin, 1984). Since the research question being raised in a meta-analysis is often dichotomous or at best one of simple relationship, the variety of more complex findings that may have appeared in the original articles is reduced. The fear has been expressed that consumers of meta-analyses may come away with nothing more than unqualified statements such as, "computer-based instruction is better than traditional lecture-based methods" or "the correlation between prior achievement and instructional support is moderately high and positive". It is certainly arguable, however, that for consumers who would not bother to digest more subtle forms of summaries, a simplistic view of the state of a research literature is better than no impression at all.

In the following sections, procedures for conducting a meta-analysis are described. Issues related to each procedure will be discussed to highlight both the potentials and the problems associated with the technique.

## A DESCRIPTION OF META-ANALYSIS

### ***Defining the Scope of the Analysis***

A first important decision to be made after a general area of research has been identified is how extensively the search will be conducted and what descriptors will be used in reviewing the literature. While this sounds like a relatively straight forward process, it is usually not. Often this step involves making literally dozens of *a priori* decisions about what will be included (and not included) in the meta-analysis. Each decision will narrow the field of search, as well as the number of studies identified and the population to which generalizations can be made.

For example, in a synthesis of mastery learning studies one might consider features such as: a) how far back in time the review will go; b) the grade level of subjects; c) the subject matter tested; d) the duration of treatment; e) whether self-paced or group-based treatments is used; e) the type and quality of the dependent variable; and f) a host of experimental design characteristics (e.g., internal and external validity). Carlberg and Walberg (1984) point out trade-offs in: a) narrowly focusing the synthesis to exclude relevant variations in treatments (high fidelity/limited conclusions); and b) making the scope of inclusion so broad that marginally relevant and/or bad research is analyzed (low fidelity/more robust conclusions).

Advice on both sides of this issue has been offered in the literature of meta-analysis. Glass, and Smith (1981) argue for the widest inclusion

criteria possible in order to reduce the effects of reviewer bias in the selection process. Eysenck (1978) has criticized this approach as "garbage in -garbage out".

As a way of accommodating this criticism, Slavin (1986, 1987) has proposed an approach called "best-evidence synthesis". This approach is based on the legal notion that "... the same evidence that would be essential in one case might be disregarded in another because in the second case there is better evidence available" (1986, p. 6). In the case of research review this means that only the best quality literature should be used in judging the general state of a research question—those studies which are high in methodological rigor and best manifest the characteristics under study. In the absence of studies of better quality, this could involve having to use less well designed studies, but in any case, comprising the best evidence. Objections to the use of this approach have been raised by Guskey (1987), who counters that the "best" in best-evidence synthesis is itself subjective and does not necessarily eliminate bias from the review.

Abrami, Cohen and d'Apollonia (1988) take a middle approach, between that advocated by Glass and that advocated by Slavin:

... we urge greater care in describing the inclusion criteria and in detailing the reasons for excluding individual studies. But we also consider that reviews sometimes go beyond describing the substance of the literature to consider the methodological problems and generalizability concerns that distinguish the best evidence from other evidence. Reviews may thus contribute to knowledge in an area through the analysis of study weaknesses as well as strengths. Such a contribution cannot be made through only the analysis of best evidence (p. 164).

### *Reviewing the Literature*

Once inclusion criteria have been established, the approach to locating studies for review is not substantially different from that used in other forms of integrative review. Primary studies may be located from a variety of sources, some of which are accessible through computer-generated searches. Most meta-analyses include the literature from relevant journals in the field. Others include theses and dissertations, conference presentations, technical reports and in-house manuscripts, chapters in books and monographs and other documents referred to categorically as "fugitive material".

Even when inclusion and exclusion criteria have been soundly determined, there remains the thorny problem of actually sorting studies by the established criteria. This process is by no means straight forward, as Abrami, Cohen and d'Apollonia (1988) have demonstrated using data from the literature on the validity of student ratings of instructors. They found that even when inclusion criteria were very clearly specified, seven expert raters had an average comprehensiveness index of only .58 (i.e., ratio of correctly included studies to incorrectly included or excluded studies) with individual indices ranging from

.13 to .88. They make recommendations for enhancing the agreement among raters, along with suggestions for improving meta-analysis methodology at four other stages in the process.

**Identifying Variables for Study**

Unless the researcher is very familiar with the primary literature under study, it is advisable to select a sample of studies for the purpose of determining variables that will be subsequently coded for analysis. The purpose of this exercise is to determine which variables, in addition to the primary distinction under study, have been most commonly reported in the literature. These additional variables may serve to aid in generalization or may actually form the basis for tests of significance in their own right. In the following sections these variables, under commonly encountered headings, are discussed.

**Demographic variables.** Among other things, these include variables related to the nature of the experimental sample under study (e.g., sex, grade level, SES).

**Treatment variables.** Included in this category are characteristics of the treatment condition, for instance, type of treatment, duration of treatment, location of treatment and experimenter characteristics.

**Design variables.** Variables falling into this category are those associated with the nature and quality of the experimental manipulation. Examples include presence of experimental control, randomization and selection, presence of pretest, nature of dependent measures, specific threats to internal and external validity

Once these variables have been identified, they are coded for each study using a scheme that is similar to that shown in Figure 1 (see page 178).

**Calculating Effect Size**

The estimate of the strength of a treatment, called an effect size, is calculated using a relatively simple procedure. For difference questions, the means of treatment and control groups are ascertained, and the control mean is subtracted from the treatment mean. Naturally, if this difference has a positive sign, it indicates that treatment subjects have outperformed control subjects, while a negative sign indicates the reverse.

This raw difference is not enough, however. It must be standardized so that other studies investigating the same variable may be averaged with it. The meta-analytic researcher accomplishes this by dividing the raw difference by an estimator of  $\sigma$ — the standard deviation of the control group (for Glass's ES), or by a pooled standard deviation (for Cohen's  $d$ ), if control and treatment standard deviations are very different. The formulae for Glass's ES and Cohen's  $d$  may be expressed as follows:

$$ES = \frac{(\text{mean}_{\text{treatment}} - \text{mean}_{\text{control}})}{\text{standard deviation}_{\text{control}}}$$

$$d = \frac{(\text{mean}_{\text{treatment}} - \text{mean}_{\text{control}})}{\text{standard deviation}_{\text{pooled}}}$$

Figure 1.  
*Cross-Sections of a Simplified Coding Scheme.*

<i>Identification</i>		
Author		
Year of Study		
<i>Demographics</i>		
Sex		1 = Male, 2 = Female
Age		1=6-9, 2-10-13, 3-14-17
<i>and others</i>		
<i>Treatment</i>		
Type of Instructions		1 = Oral, 2 = Written, 3 = Unspecified
Duration of Treatment		1 = <1 hour, 2 = >1 hour to 1 day, 3 = >1 day to 1 week
<i>and others</i>		
<i>Design</i>		
Random Assignment		1 = Complete, 2 = Stratified, 3 = Blocked, 4=None
<b>Mortality</b>		1=0, 2=<5%, 3=>5%<10%, 4=Unspecified
<i>and others.</i>		

Other formulae for deriving effect sizes in studies that do not contain some of the elements listed above have been presented by McGaw and Glass (1980). Formulae are also available for obtaining effect sizes when transformed scales are used (e.g., gain scores), when factorial designs are used or when dependent measures have been adjusted by a covariate.

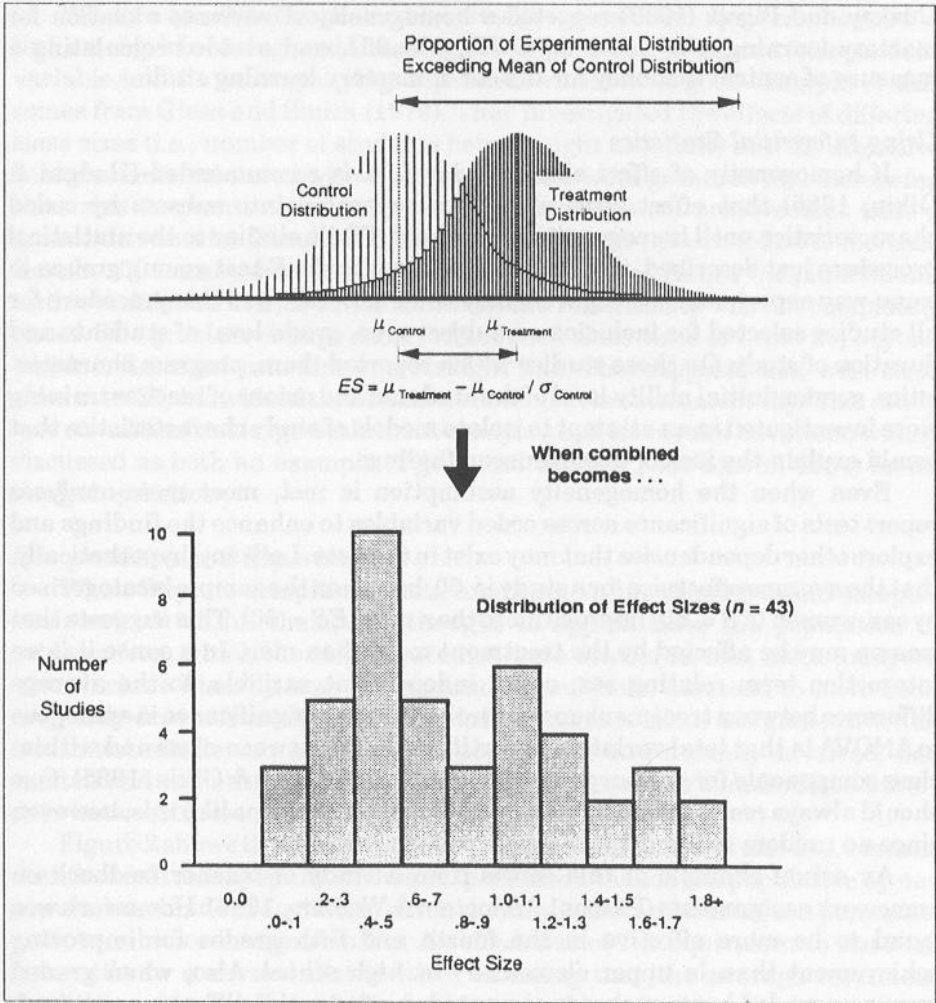
The result is a z-score of sorts\* - a standardized metric which represents the number of standard deviations the treatment condition has outperformed the control condition (or underperformed if the sign is negative). All of the effect sizes in the study are then averaged (to produce a mean effect size) or the median of the distribution is represented. Figure 2 (see page 179) shows how the difference between the two theoretical distributions of control and treatment may be shown graphically, and then represented as an actual distribution of effect sizes.

---

**\*Z-scores are calculated within a distribution of raw scores using the following formula: Score in the Distribution / (Distribution Mean + Distribution Standard Deviation). Since the distribution of z-scores is in unsquared deviation units, its mean is always 0 and its standard deviation is always 1.0. This is not the case with an effect size distribution, since the mean and standard deviation for each study included is different. The mean of the distribution may be either positive or negative and represents the average standardized difference between sample**



Figure 2.  
*Distribution of Experimental and Control Differences and the Resulting Effect Size Distribution.*



Note: Effect size distribution from Guskey and Pigott (1988).

At this stage, the homogeneity of the effect size distribution is considered. Homogeneity of effect size is the assumption that the variance surrounding an effect size mean is small enough for the mean to be considered a valid representation of the phenomenon in the population, and is similar to the assumption of homogeneity of variance in experimental design. The test is accomplished by comparing a calculated sum of squares to the chi-square distribution (with  $k$  degrees of freedom where  $k$  is the total number of studies).

In some meta-analyses that are published, this test is either not performed or not reported, although it is critical to interpreting the average effect size.

The finding of an effect size of .94 with a standard deviation of 1.91 in a mastery learning meta-analysis (Lysakowski & Walberg, 1982) is probably an example of too much variability, considering the magnitude of the mean. Consequently, Guskey and Pigott (1988) reported a homogeneity of variance violation for mastery learning,  $\chi^2 = 759.5$  ( $df = 77$ ),  $p < .001$ , and avoided calculating a measure of central tendency for the set of mastery learning studies.

### *Using Inferential Statistics*

If homogeneity of effect size is violated, it is recommended (Hedges & Olkin, 1985) that effect sizes should be separated into subsets by coded characteristics until homogeneity is achieved. While similar to the statistical procedure just described, this test is equivalent to the F-test among groups in a one-way experimental design. Guskey and Pigott followed this procedure for all studies selected for inclusion on subject area, grade level of students and duration of study. On those studies which reported them, program characteristics, gender, initial ability level of the students and extent of teacher training were investigated in an attempt to isolate models of study characteristics that would explain the lack of homogeneous findings.

Even when the homogeneity assumption is met, most meta-analyses report tests of significance across coded variables to enhance the findings and explore other dependencies that may exist in the data. Let's say, hypothetically, that the average effect size for a study is .60, but when the sample is categorized by sex, women ( $ES = .80$ ) improve more than men ( $ES = .60$ ). This suggests that women may be affected by the treatment more than men. In a sense it is an interaction term relating sex, as an independent variable, to the average difference between treatment and control. The test of significance is analogous to ANOVA in that total variation is partitioned into between-class and within-class components for the purposes of comparison (Hedges & Olkin, 1985). One should always resist a causal interpretation of comparisons like this, however, since no random assignment to treatments is involved.

An actual example of this comes from a study of teacher feedback on homework assignments (Paschal, Weinstein & Walberg, 1984). Homework was found to be more effective in the fourth and fifth grades for improving achievement than in upper elementary or high school. Also, when graded versus ungraded homework was compared, a substantial difference emerged. Graded homework produced an effect size of .80, while ungraded homework influenced achievement by only .36 standard deviations. Both of these characteristics of the sample produced significant differences when tested using the procedures outlined above.

In addition, researchers would be interested in whether there is a difference among a variety of methodological and demographic aspects of the studies under consideration. This amounts to searching for bias in the variables coded under threats to internal and external validity, publication sources, such as articles, book chapters, dissertations and ERIC documents and other variables that may reside concomitantly with treatments. Not surprisingly, higher effect

sizes are often found for published over non-published works, since journals usually accept studies that apply more rigorous methods and often reject studies reporting no significant differences.

Where a quantitative scale is involved, regression analysis can be used to test among increasing or decreasing levels of some continuous independent variable and its accompanying dependent variable. A good example of this comes from Glass and Smith (1979). They investigated the effects of differing class sizes (i.e., number of students being taught at a time) and the cognitive achievement associated with it. Eighty studies were gathered and increasing class size (a quantitative scale) was regressed against achievement (also a quantitative scale). Results indicated that achievement was found to increase from by .50 standard deviations as class size changed from 1 (i.e., individualized tutorial instruction) to 40. However, the relationship was not completely linear. The greatest change occurred between class sizes of 1 and 20, beyond which it flattened into almost a straight line. This suggests that with class sizes over 20, individual achievement does not rise incrementally. This study was one of the first large-scale meta-analyses and its results have been widely discussed as both an example of good and bad (e.g., see Slavin, 1984) meta-analytic practice.

### ***Interpretation and Reporting***

When they are completed, meta-analyses, unlike the individual samples summarized within them, are thought to approximate the population of subjects from which the original studies were drawn. In fact, meta-analyses often include literally tens of thousands of subjects, assumed to have been originally drawn from the same population before random assignment. When treatment effects are present, two populations are actually involved, one treated and one untreated. The effect size estimates the standardized difference between these populations.

Figure 2 shows this comparison in graphic form. An effect size of 1.0 means that the treatment population has outperformed the control population by one population standard deviation. Often the effect size is converted to a percentile rank to enhance interpretability. An effect size of 1 is equivalent to the 84th percentile in a normal distribution, meaning that the average treatment condition subject is above 84% of subjects in the control condition.

Since one of the purposes of meta-analytic studies is to allow for comparison among potentially useful instructional treatments, some additional form of interpretation of average effect size is desirable. A non-technical interpretation of low, medium and high effect sizes has been suggested by Cohen (1969). Small effect sizes (e.g., .20 or the 58th percentile) are similar to those associated with comparisons among the heights of 15 and 16 year old girls. Medium effect sizes (e.g., .50 or the 69th percentile) would be similar to differences between 14 and 18 year old girls. Large effect sizes (.80 or the 79th percentile) are of the order of magnitude of differences in IQ between holders of Ph.D. degrees and the average college freshman.

## THE EXAMPLE OF MASTERY LEARNING

### ***Description of Mastery Learning***

Although the concept of mastery learning has existed since the 1920s, it became a mainstream instructional strategy primarily as a result of work by Bloom (1968) and Block and Anderson (1975). In its simplest form mastery learning is "... a test about what the student was supposed to learn; a test not for grading or judging, but rather to see what the student has learned and what he or she needs to learn. The students are then given some help" (Bloom quoted in Koerner, 1986, p. 60). There are two primary forms that have grown out of this basic notion: a) group-based mastery learning; and b) personalized system of instruction (PSI/Keller Plan). PSI is an individualized form of mastery learning.

Supporters of mastery learning claim that the method will produce significantly higher achievement results, given the same objectives, the same materials, and the same amount of time allocation as standard instructional models. In group-based mastery instruction, teachers determine the pace of instruction, while in PSI the student controls the pace. In addition, it is argued that learning achievement will be dramatic: 90% of the learners will be able to achieve at a learning level of 85% or higher. Effectively, this would change the normal distribution of learning outcomes produced by standard instruction into one that is highly negatively skewed (Bloom, 1984). Guidance is individualized and focused on what has not been achieved.

Over the years many studies have been conducted to test these claims in both the contexts of group-based settings and PSI. The first review of literature (Block & Burns, 1976), conducted on both group-based and PSI studies, concluded that the mastery approaches described result in higher achievement and positive affective outcomes. However, the cognitive results were not as dramatic as the supporters of mastery learning had claimed. In the 1980s three meta-analytic studies of group-based mastery learning were conducted, each reaching dramatically different results as to the state of research that underlies this instructional technique and the magnitude of treatment effects. These studies are summarized in Table 1 (see pages 183 and 184), and their features and issues related to them are discussed.

### ***Three Meta-Analyses on Mastery Learning***

The first major review of research (Lysakowski and Walberg, 1982) was a quantitative summary of three of the four fundamental ingredients of quality instruction: cues, participation and feedback corrective. Reinforcement, the fourth element, had been reviewed previously. The reviewers concluded that the average effect for all three components was .97, and that the effect for feedback and correction, the element most commonly associated with mastery learning, was .94. Clearly, this was dramatic evidence that mastery learning had achieved the potential that had all along been claimed for it.

Five years later, Slavin (1987) published a meta-analytic study of mastery

TABLE 1  
Comparison of Three Meta-Analytic Studies of Group-Based Mastery Learning

Features	Lysakowski & Walberg (1982)	Slavin (1987)	Guskey & Pigott (1988)
<i>Purpose</i>	Effect size estimation for overall literature on mastery learning	Effect size estimation of three theoretical claims	Effect size estimation for mastery learning applications.
<i>Grade Levels</i>	Preschool through adult settings	Elementary and secondary	Elementary and college
<i>Measures Considered</i>	Cognitive and affective	Cognitive, standardized and locally produced	Locally produced cognitive and affective
<i>Selection Criteria</i>	All quantitative studies: a) in an educational setting; b) in normal school-type setting; and c) where contrasted groups or correlational groups were included.	Quantitative studies: a) germaneness to definition of mastery learning; and b) rigorous methodological adequacy (i.e., experimental design)	All studies that: a) reported data on measured outcomes for all conditions and had a clear time-series design; and b) were free from serious methodological flaws.
<i>Exclusion Criteria</i>	Studies of: a) subjects with special needs; b) computer-assisted and programmed instruction; and c) contrived studies out of real classroom.	Studies where grades and time on-task were the only dependent measures	Theoretical or descriptive articles that contained no quantitative analyses.

TABLE 1 (continued)  
Comparison of Three Meta-Analytic Studies of Group-Based Mastery Learning

Features	Lysakowski & Walberg (1982)	Slavin (1987)	Guskey & Pigott (1988)
Number of Studies Selected	20	7 9 4	46
Percentage of Literature Overlap	With Guskey (2/20) 10% With Slavin (0/20) 0%	With Lysak..(6/16) 38% With Guskey (0/16) 0%	With Lysak..(2/46) 4% With Slavin (6/46) 13%
Effect Size(s)	Mean = .94	With Lysak..(6/16) 38% With Guskey (0/16) 0%	With Lysak..(2/46) 4% With Slavin (6/46) 13%
Variability	SD = 1.91	None indicated	None indicated
		"SC" <sup>1</sup> "CFC" <sup>2</sup> "ETC" <sup>3</sup>	"SC" <sup>1</sup> "CFC" <sup>2</sup> "ETC" <sup>3</sup>
		Mean = .04 Median = .26 Median = .31	Mean undetermined because of heterogeneity of studies. Median interval = .6 -.7

<sup>1</sup> "Strong Claim" comes from studies where mastery and control groups have same objectives, same materials and same amount of time with standardized measures.

<sup>2</sup> "Curricula Focus Claim" (weak claim) comes from studies where teachers use a particular curricula and students learn a particular set of objectives.

<sup>3</sup> "Extra Time Claim" (weak claim) comes from studies where additional time and instructional resources were required to bring all students to an acceptable level of achievement.

learning that all but refuted the major claims made by Bloom and others. Using a technique developed by him, called "best-evidence synthesis," he was able to show that the results of mastery learning are considerably smaller in subsets of studies embodying: a) the "strong claim"- that mastery will outperform the control group when they have the same objectives, the same materials and the same amount of time and when learning is measured with standardized instruments; b) the "curricular focus claim" — that mastery learning focuses teachers on particular curricula and students on the attainment of particular objectives; and c) the "extra time claim"- that mastery learning is an effective use of additional time and instructional resources to bring all students to an acceptable level of achievement. In addition, Slavin only used studies that he considered methodologically rigorous and those where the mastery learning treatment lasted for four weeks or longer. Evidence for the "strong claim" produced a median effect size of .04 (essentially 0). Studies representing the "curricula focus claim" were found to have a median effect size of .26, and the median for those representing the "extra time claim" was .31.

The most recent meta-analysis of mastery learning studies was conducted by Guskey & Pigott (1988). A subset of articles related only to elementary and secondary classrooms had been published previously by Guskey and Gates (1986). The Guskey and Pigott meta-analysis included a larger number of studies ( $n = 46$ ) and arrived at a somewhat surprising conclusion: that the variability of effect sizes for methodologically sound studies of group-based mastery learning was too great to compute an average effect size estimate. Attempts to derive models of measured variables which explained this heterogeneity were generally unsuccessful, although some trends are noted (e.g., higher in some subject matters).

### ***Some Reasons for the Differences***

There are several explanations for the discrepancies among these meta-analyses which help to demonstrate some of the characteristics of meta-analyses in general. First, it is obvious that meta-analyses conducted at different points in time, especially when there is high research productivity in the area, are bound to produce different results. More refined methods of study, better research designs, sensitivity to criticisms of previous research studies and a host of other considerations can affect the results achieved through meta-analysis from one era to another.

Second, the selection of studies for inclusion, even when the same study pool is available, can dramatically affect the results that are achieved by different researchers. Smith and Glass (1977) and others argue for the inclusion of all available studies that include the minimum criterion of a metric of comparison, while others have claimed that mixing apples and oranges clouds the issue under study considerably, rather than elucidating it (Slavin, 1984). In the meta-analyses under consideration, the earliest effort set wide inclusion criteria that admitted many studies that were not included in the subsequent articles. Note in Table 1 that the percentage of overlap between

Lysakowski/Walberg and Slavin and Guskey/Pigott is 0% and 10%, respectively. Between Slavin and Guskey/Pigott the overlap is considerably higher. In this latter case, however, it is far less than it could be because of the restrictive conditions set by the "best evidence synthesis".

One interesting finding by Guskey and Pigott (and Lysakowski & Walberg, although it was not discussed) reveals something of interest about the outcomes of meta-analytic studies. You may notice in Table 1 that for an effect size of .94, Lysakowski & Walberg report a standard deviation of 1.91. This describes a very platykurtic distribution (i.e., flat) which cannot be thought of as homogeneously summarized by a single effect size. In sampling an essentially different literature of mastery learning studies, Guskey and Pigott found the same thing. The standard deviation of the distribution of effect sizes should be considered an important piece of information and effect sizes with high standard deviations or standard error of the mean should not be taken at face value, because they may not be statistically significant.

The variability, both within and among mastery learning studies probably says little about the pedagogical soundness of its best applications, but instead bespeaks the implementation and methodological problems that continue to plague it. Variations in practice abound and there exists thorny research design problems which have not yet been fully addressed. These include dealing with time to mastery, equilibrating mastery and non-mastery treatments, dealing with the skewed distributions that invariably results from good mastery applications and establishing a sound rationale for using either standardized or well constructed locally produced instruments. As the technology of mastery learning improves and researchers become sensitive to methodological problems that are peculiar to mastery investigation, the variability in mastery studies will undoubtedly subside.

#### GUIDANCE FOR INTERPRETATION AND USE OF META-ANALYSIS

##### ***Comments for Practitioners and Researchers***

Conscientious practitioners are always searching for support for the design of quality instructional programs. This might come from previous successes, from the analysis of cost/benefits, or from the literature of research studies. Meta-analysis seems a reasonable tool for achieving the latter goal.

Table 2 (see pages 187 through 191) lists 26 meta-analyses of instructional variables divided for convenience into categories: instructional media, text design features, classroom processes, feedback and correction and social aspects of learning. The references to these studies are provided in the appendix. These studies represent a potentially valuable resource for the practitioner and researcher alike.

In spite of the apparent flaws in the practice of meta-analysis, it remains the single most powerful tool for summarizing studies in an era of rapidly



TABLE 2  
Summary of Meta-Analytic Studies of Instructional Variable Effects on Student Achievement

Author(s)	Year	Instructional Variable	Number of Studies	Mean Effect Size <sup>1</sup> (SD or SE)	Other Selected Findings
<i>Instructional Media</i>					
Kulik, Kulik & Cohen	1979a	Audio-Tutorial	48	$ES = .20, d = .20$	•Journal sources > other sources
Kulik, Kulik & Cohen	1980	Computer-Based College Teaching	54	$d = .25 (SD = .64)$	•Instr. Time: CBI < conventional
Kulik, Cohen & Ebeling	1980	Programmed Instruction in Higher Education	56	$ES = .24 (SD = .45)$	•More recent studies > older •N.S.D. attitudes or withdrawal
Cohen, Ebeling & Kulik	1981	Visual-Based <sup>2</sup> Instruction	65	$d = .15 (SD = .44)$	•Doctoral institutions > others •N.S.D. attitudes or withdrawal
Kulik, Shwalb & Kulik	1982	Programmed Instruction in Secondary School	47	$ES = .08 (SD = .48)$	•Attitudes PI < other

TABLE 2 (continued)  
 Summary of Meta-Analytic Studies of Instructional variable Effects on Student Achievement

Author(s)	Year	Instructional Variable	Number of Studies	Mean Effect Size <sup>1</sup> (SD or SE)	Other Selected Findings
Kulik, Bangert & Williams	1983	Computer-Based in Secondary School	48	ES=.32 (SD=.42)	<ul style="list-style-type: none"> <li>•Attitudes CBI &gt; others</li> <li>•Short duration &gt; longer duration</li> </ul>
Kulik, Kulik & Shwalb	1986	Computer-Based Adult Education	23	ES=.42 (SE=.11)	<ul style="list-style-type: none"> <li>•Instr. Time: CBE &lt; conventional</li> <li>•Comp. Enriched Inst. &gt; CMI &amp; CAI</li> </ul>
Shwalb, Shwalb Azuma	1986	Educ. Tech. in Japanese Schools <sup>3</sup>	116	ES=.41 (SD=.27)	<ul style="list-style-type: none"> <li>•PI most effective &amp; technology</li> </ul>
Nierniec, Samson, Weinstein & Walberg	1987	Computer-Based Elementary School Instruction	48	ES=.45 (SD=.36)	<ul style="list-style-type: none"> <li>•Many demographics</li> </ul>
<b>Text Design Features</b>					
Luiten, Ames & Ackerson	1980	Advance Organizers	135	ES=.21 (SE=.04)	<ul style="list-style-type: none"> <li>•21 day delay ES=.30, +22 ES=.38</li> <li>•Aural presentation ES=.37</li> </ul>

TABLE 2 (continued)  
Summary of Meta-Analytic Studies of Instructional Variable Effects on Student Achievement

Author(s)	Year	Instructional Variable	Number of Studies	Mean Effect Size <sup>1</sup> (SD or SE)	Other Selected Findings
Stone	1983	Advance Organizers	112	ES=.66 (SD =.78) Median = .48	•Conceptual > concrete •Highest for medium ability
Levie & Lentz	1982	Illustrations <sup>4</sup>	46	.50	•None indicated
Klauer	1984	Beh. Obs., learning directions or questions	23	Intentional ES=.40 Incidental Learning ES= -.20	•None indicated
Moore & Readence Comprehen.	1984	Graphic Organizers	23	ES=.22 (SD=.58)	•Vocabulary lring. >
Hamaker	1986	Adjunct Prequestions Adjunct Postquestions	23 59	ES=.15 (SD=.15) <sup>5</sup> ES=.16 (SD=.10)	•Repeated dv questions > others
<b>Classroom Processes</b>					
Lysakowski & Walberg	1981	Classroom Reinforcement	39	ES=1.17 (SD=1.0)	•Tangible > intangible
Redfield & Rousseau	1981	Higher-Order Questions	14	ES=.73	•None indicated

TABLE 2 (continued)  
 Summary of Meta-Analytic Studies of Instructional variable Effects on Student Achievement

Author(s)	Year	Instructional Variable	Number of Studies	Mean Effect Size <sup>1</sup> (SD or SE)	Other Selected Findings
Lysakowski & Walberg	1982	Teaching Cues	17	ES=1.28 (SD=1.15)	•None indicated
Lysakowski & Walberg	1982	Student Participation	22	ES=.88 (SD=.94)	•None indicated
<i>Feedback and Correction</i>					
Kulik, Kulik & Cohen	1979 <sup>5</sup>	Personalized System of Instruction (PSI)	20	ES=.49	•Course grades ES=.69
Lysakowski & Walberg	1982	Corrective Feedback	20	ES=.94 (SD=1.91)	•None indicated
Cohen, Kulik & Kulik classrooms	1982	Student Tutoring	52	ES=.40 (SE=.07)	•Attitudes > in tutored
Paschal, Weinstein & Walberg	1984	Homework (overall) Graded only	15	ES=.36 (SD=.24)	•Homework > for 4th & 5th graders
			7	ES=.80 (SD=.29)	
Guskey & Gates	1986	Mastery Learning	27	Undetermined <sup>6</sup>	•ES > in elementary & jr. high

TABLE 2 (continued)  
 Summary of Meta-Analytic Studies of Instructional Variable Effects on Student Achievement

Author(s)	Year	Instructional Variable	Number of Studies	Mean Effect Size <sup>1</sup> (SD or SE)	Other Selected Findings
Slavin	1987	Mastery Learning Elementary and Secondary "Best-evidence Synthesis"	7 <sup>7</sup> 8 <sup>8</sup> 4 <sup>9</sup>	Median=.04 Median=.26 Median=.31	•None indicated
Guskey & Pigott	1988	Mastery Learning All Grade Levels	46	Undetermined <sup>6</sup> Median=.60-.70	•No underlying demographic model found to explain heterogeneity
<b>Social Aspects of Learning</b>					
Ide, Parkerson, Haertel & Walberg	1981	Peer Group Influence	10	$r = .24^{10}$ , $r^2 = .06$	•Higher in urban settings

<sup>1</sup> ES = Glass's ES ( $M_s - M_c / SD_s$ );  $d$  = Cohen's  $d$  ( $M_s - M_c / S_p$  pooled) or by  $r$  = Pearson Product-Moment Correlation. All effect sizes are from achievement data. *SD* refers to standard deviation and *SE* refers to the standard error of the mean. Unless otherwise indicated, effect sizes shown are for overall research question only.

<sup>2</sup> Included were still projection, film, multimedia, CCTV, ETV, Observation and feedback.

<sup>3</sup> Technologies reported were programmed instruction, visually-based instruction, multi-media, CAI and audio-tutorial.

<sup>4</sup> Technically not a meta-analysis. ES is based on conversion of percentage of increase of treatment over control to standardized units.

<sup>5</sup> Reported for studies where repeated questions (either pre- or post-) were used with repeated criterion questions.

<sup>6</sup> ESs for all studies reviewed were considered too heterogeneous to derive a mean effect size.

<sup>7</sup> "Strong Claim" (see Mastery Learning example).

<sup>8</sup> "Curriculum Focus Claim" (see Mastery Learning example).

<sup>9</sup> "Extra Time Claim" (see Mastery Learning example).

<sup>10</sup> Median correlation where values range from .67 to -.16 on all outcome measures.

expanding scientific literature. This process is necessary even when flawed because of the impossibility of keeping up with even a fraction of the literature by researchers and practitioners alike.

For the researcher, meta-analysis represents a means for focusing thought on the large questions and a heuristic for designing future studies taking into account the smaller questions. For the practitioner in the media and technology field, meta-analysis is a means for making broad decisions about the implementation of new programs and the design of instructional products. Effect size is the metric for predicting what might happen in a new circumstance if a particular instructional variable were implemented. It tells roughly how many standard deviations of additional achievement would be expected over groups that do not receive the variable. However, it behooves both practitioners and researchers alike to heed the warnings of the critics of meta-analysis practice. The following suggestions may aid the reader in using the information contained in meta-analysis to support their instructional decisions.

1. *Achieving a common definition* — While seemingly self-evident, the consumer of meta-analyses should make certain that their definition and that of the author are in agreement and that the studies reported in the meta-analysis are examples of the conceptual definition under consideration. For instance, a designer searching for pre-instructional activities for textbook design should realize that the studies reported under the rubric of advance organizers will not include other design features that might be commonly associated with advance organization, such as outlines, abstracts, introductions and overviews. The technical definition created by Ausubel and tested in meta-analyses does not include the above.

2. *Achieving a common circumstance* — Meta-analyses often summarize studies across a wide variety of instructional or educational circumstances (i.e., grade levels, SES levels, geographical boundaries). Consumers of meta-analyses should be aware of these circumstances and if necessary base their conclusions on subsets within the meta-analysis that fit their own needs. There is a danger in this, however. When the studies in a meta-analysis are subdivided, the resulting number of studies per subset is often quite small, often fewer than 5 studies. It is more difficult to base a firm judgment on smaller data set than larger ones, because the smaller number runs a greater risk of a Type II error (accepting a when it should be rejected). Naturally, the variability among studies within a subset should be of concern, as well as the mean.

3. *Achieving an overall understanding* — Of note in Table 2 is the fact that some areas have been investigated several times. This is partly because the state of evidence is always advancing. More recent meta-analyses, supplant older ones in characterizing the field more fully. However, in some cases, a meta-analysis may be repeated to reconsider an earlier finding or to incorporate a new methodological or conceptual application into the state of the art. The meta-analysis of the mastery learning literature by Slavin (1987) is a good example of the latter case. The "best-evidence synthesis" represents a new

conception of how inclusion and exclusion criteria should be set and as a result a dramatically different impression of the field emerged.

It is surprising that follow-ups of the Kulik et al. media studies of the early 1980s have not been attempted, particularly after Clark's 1983 attack on their validity. One would have expected a response to determine if Clark's assertion accurately represented the overall literature, given that his findings were based on only a partial selection of these studies. We can only speculate that calls by Salomon and Clark (1979) and Bernard (1986) and others to stop asking gross media questions, comparing a media treatment to a control group, have been heeded. Unfortunately, literature concerning the nuances of within media comparisons does not abound, reducing the likelihood of additional meta-analyses in the media area.

Since new meta-analyses can appear for either of the reasons mentioned above, to achieve a complete understanding of a field of inquiry, it is important to become familiar with all of the meta-analyses that have been conducted, not just the most recent ones.

Another point of importance here is the limitations of meta-analysis for drawing specific conclusions about when or under what exact circumstances a particular technique or medium should be applied. Meta-analysis is far too global to aid in the fine-grained analysis of instructional problems. In addition, it has seldom been used to address instructional treatments that are continuous or incrementally applied (e.g., varying degrees of feedback) or where variations in type of common strategy (e.g., type of questioning) are examined. Therefore, meta-analysis is most useful as a tool for making the larger instructional decisions. The designer must look to more specific studies of instructional treatments and/or conduct local evaluation studies on prototype materials in order to gain insight into particular aspects of developing instruction.

**4. Achieving a statistical understanding** — There are several important points here. One, the mean effect size that is reported may not accurately reflect the underlying population parameter. If a test of homogeneity of effect size is not provided, look carefully at the magnitude of the standard deviation if it is given. Interpretation of this statistic may be supplemented by a histogram similar to the one pictured at the bottom of Figure 1 that are often included (stem-and-leaf diagrams are also common). This will provide a visual sense of the distribution of effect sizes and the variability among them. Two, while tests of significance within the distribution of effect sizes and between subsets of demographics are important, they can be misleading. When sample size is small, the power of the test is low reducing the probability that differences will be detected, even when they are present in the population. When sample size is large, even a relatively small effect size may exceed the critical value necessary to reject the null hypothesis. Three, in interpreting the effect size, reference to the percentile rank and to non-technical descriptions of the meaning of effect sizes are invaluable.

*Comments for Researchers*

Meta-analysis represents both a means for estimating the effects of instructional treatments in practice and a heuristic for designing future research studies. This latter function may be accomplished in two ways. First, meta-analyses can sensitize researchers to issues of design and methodology that will admit or not admit their studies to scrutiny in future attempts to synthesize the literature. Second, the ancillary analyses contained in most meta-analyses can aid researchers in identifying the sources of data and variables that are likely to interact with the major question that is being addressed. If these suggestions sound like prescriptions for conformity, that is exactly how they are intended. Progress in the science of instruction, to some degree, is predicated on the presence of high quality replications in order for the larger questions to be answered.

However, it should be recognized that there exist limitations to meta-analysis as a heuristic for research. Meta-analysis is a retrospective approach which derives its strength from the weight of past efforts. It is therefore unlikely that new developments—those that will qualitatively extend beyond present practice—will emerge from this technique. Meta-analysis will never be a substitute for insight and creativity in the conduct of primary research or the development of new instructional methods. In short, as a technology of quantitative synthesis, meta-analysis should never substitute for the kind of in-depth exploration and complex thinking that characterize productive scientific enquiry.

## CONCLUSION

In this article we have sketched a broad picture of the nature of meta-analysis, its potential for informing researchers about the overall effectiveness of variables in a given field and for aiding media and technology practitioners in making decisions concerning larger instructional development issues. We have discussed both the philosophical and practical objections to meta-analysis and have described the process of doing a meta-analysis in some detail. Clearly, all of the many issues that have arisen over the last 15 years cannot be catalogued here. However, the core issues that have been represented and the references, provide ample fodder for further consideration.

## REFERENCES

- Abrami, P. C., Cohen, P. A., & d'Apollonia, S. (1988). Implementation problems in meta-analysis. *Review of Educational Research*, 58(2), 151-179.
- Bernard, R. M. (1986). Is research in new technology caught in the same old trap? *Canadian Journal of Educational Communication*, 147-151.



- Block, J. H. (1971). Introduction to mastery learning: Theory and practice. In J. H. Block (Ed.) *Mastery learning: Theory and practice* (p. 2-12). New York: Holt, Rinehart & Winston.
- Block, J. H., & Anderson, L. W. (1975). *Mastery learning in classroom instruction*. New York: McMillian.
- Block, J. H., & Burns, R. B. (1976). Mastery learning. In L. S. Schulman (Ed.), *Review of Research in Education* (Vol. 4, pp. 3-49). Itasca, IL: F. E. Peacock.
- Bloom, B. S. (1968). Learning for mastery (UCLA-CSEIP) *Evaluation Comment*, 1(2), 1-12.
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4-16.
- Carlberg, C. G., & Walberg, H. J. (1984). Techniques of research synthesis. *Journal of Special Education*, 18(1), 11-26.
- Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53(4), 445-459.
- Clark, R. E. (1985a). Confounding in educational computing research. *Journal of Educational Computing Research*, 1, 137-148.
- Clark, R. E. (1985b). Evidence for confounding in computer-based instruction studies: Analyzing the meta-analyses. *Educational Communication and Technology Journal*, 33(4), 249-262.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist*, 43, 517.
- Eysenck, H. J. (1984). Meta-analysis: An abuse of research integration. *Journal of Special Education*, 18(1), 41-59.
- Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher*, 5(10), 3-8.
- Glass, G. V. (1978). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 361-379.
- Glass, G. V., & Smith, M. L. (1979). Meta-analysis of the research on class size and achievement. *Educational Evaluation and Policy Analysis*, 1, 2-16.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage
- Guba, E. G. (1979). Naturalistic inquiry. *Improving Human Performance Quarterly*, 8, 268-276.
- Guskey, T. R. (1987). Rethinking "mastery learning reconsidered". *Review of Educational Research*, 57(2), 225-229.
- Guskey, T. R., & Gates, S. L. (1986). Synthesis of research on the effects of mastery learning in elementary and secondary schools. *Educational Leadership*, 43(8), 73-80.
- Guskey, T. R., & Pigott, T. D. (1988). Research on group-based mastery learning programs: A meta-analysis. *Journal of Educational Research*, 81(4), 197-216.

- Guskin, S. L. (1984). Problems and promises of meta-analysis in special education. *Journal of Special Education*, 18(1), 73-80.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Jackson, G. B. (1980). Methods for integrative reviews. *Review of Educational Research*, 50, 438-460.
- Kavale, K. A. (1984). Potential advantages of the meta-analysis technique for research in special education. *Journal of Special Education*, 18(1), 61-92.
- Koerner, T. F. (1986, November). A discussion about instruction and learning, teachers and schools. *NASSP Bulletin*, 70, 63-69.
- Light, R. J., & Smith, P.V. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard Educational Review*, 41, 429-471.
- Lysakowski, R. S., & Walberg, H. J., (1982). Instructional effects of cues participation, and corrective feedback: A quantitative synthesis. *American Educational Research Journal*, 19(4), 559-578.
- McGaw, B., & Glass, G. V. (1980). Choice of the metric for effect size in meta-analysis. *American Educational Research Journal*, 17(3), 326-337.
- Paschal, R. A., Weinstein, T., & Walberg, H. J. (1984). The effects of homework on learning: A quantitative synthesis. *Journal of Educational Research*, 78(2), 97-104.
- Petkovitch, M. D., & Tennyson, R. D. (1984). Clark's "learning from media": A critique. *Educational Communication and Technology Journal*, 32(4), 233-241.
- Saettler, P. (1968). *A history of instructional technology*. New York: McGraw-Hill.
- Salomon, G., & Clark, R. E. (1979). Re-examining the methodology of research on media and technology in education. *Review of Educational Research*, 47(1), 99-120.
- Slavin, R. E. (1984). Meta-analysis in education: How has it been used? *Educational Researcher*, 13(8), 6-15, 24-27.
- Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, 15(9), 5-11.
- Slavin, R. E. (1987). Mastery learning reconsidered. *Review of Educational Research*, 57(2), 175-213.
- Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.

## APPENDIX

*References to Meta-Analyses in Table 2*

- Cohen, P. A., Ebeling, B. J., & Kulik, J. A. (1981). A meta-analysis of outcome studies of visual-based instruction. *Educational Communication and Technology Journal*, 29(1), 26-36.

- Cohen, P. A., Kulik, J. A., & Kulik, C.-L. C., (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19(2), 237-248.
- Guskey, T. R., & Gates, S. L. (1986). Synthesis of research on the effects of mastery learning in elementary and secondary schools. *Educational Leadership*, 43(8), 73-80.
- Guskey, T. R., & Pigott, T. D. (1988). Research on group-based mastery learning programs: A meta-analysis. *Journal of Educational Research*, 81(4), 197-216.
- Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research*, 56(2), 212-242.
- Ide, K. J., Pakerson, J., Haertel, G. D., & Walberg, H. J. (1981). Peer group influence on educational outcomes: A quantitative synthesis. *Journal of Educational Psychology*, 73(4), 472-484.
- Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, 89(1), 47-62.
- Klauer, K. J., (1984). Intentional and incidental learning with instructional texts: A meta-analysis of 1970-1980. *American Educational Research Journal*, 21(2), 323-339.
- Kulik, C.-L. C., Kulik, J. A., & Shwalb, B. J. (1986). The effectiveness of computer-based adult education: A meta-analysis. *Journal of Educational Computing Research*, 2(2), 235-251.
- Kulik, C.-L. C., Shwalb, B. J., & Kulik, J. A. (1982). Programmed instruction in secondary education: A meta-analysis of evaluation findings. *Journal of Educational Research*, 75(3), 133-138.
- Kulik, J. A., Bangert, R. L., & Williams, G. W. (1983). Effects of computer-based teaching on secondary school students. *Journal of Educational Psychology*, 53, 143-158.
- Kulik, J. A., Cohen, P. A., & Ebeling, B. J. (1980). Effectiveness of programmed instruction in higher education: A meta-analysis of findings. *Educational Evaluation and Policy Analysis*, 2(6), 51-64.
- Kulik, J. A., Kulik, C.-L. C., & Cohen, P. A. (1979a). Research on audio-tutorial instruction: A meta-analysis of comparative studies. *Research in Higher Education*, 11(4), 321-341.
- Kulik, J. A., Kulik, C.-L. C., & Cohen, P. A. (1979b). A meta-analysis of outcome studies of Keller's Personalized System of Instruction. *American Psychologist*, 34(4), 307-318.
- Kulik, J. A., Kulik, C.-L. C., & Cohen, P. A. (1980). Effectiveness of computer-based college teaching: A meta-analysis of findings. *Review of Educational Research*, 50(4), 525-544.
- Levie, H. W., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology Journal*, 30(4), 195-232.

- Luiten, J., Ames, W., & Ackerson, G. (1980). A meta-analysis of the effects of advance organizers on learning and retention. *American Educational Research Journal*, 17(2), 211-218.
- Lysakowski, R. S., & Walberg, H. J. (1981). Classroom reinforcement and learning: A quantitative synthesis. *Journal of Educational Research*, 75(2), 69-77.
- Lysakowski, R. S., & Walberg, H. J., (1982). Instructional effects of cues, participation, and corrective feedback: A quantitative synthesis. *American Educational Research Journal*, 19(4), 559-578.
- Moore, D. W., & Readance, J. E. (1984). A quantitative and qualitative review of graphic organizer research. *Journal of Educational Research*, 78(1), 11-17.
- Niemiec, R., Samson, G., Weinstein, T., & Walberg, H. J. (1987). The effects of computer based instruction in elementary schools: A quantitative synthesis. *Journal of Research on Computing in Education*, 19(2), 85-103.
- Paschal, R. A., Weinstein, T., & Walberg, H. J. (1984). The effects of homework on learning: A quantitative synthesis. *Journal of Educational Research*, 78(2), 97-104.
- Redfield, D. L., & Rousseau, E. W. (1981). Meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research*, 51(2), 235-245.
- Shawlb, B. J., Shawlb, D. W., & Azuma, H. (1986). Educational technology in the Japanese schools - A meta-analysis of findings. *Educational Technology Research*, 9, 13-30.
- Slavin, R. E. (1987). Mastery learning reconsidered. *Review of Educational Research*, 57(2), 175-213.
- Stone, C. L. (1983). A meta-analysis of advance organizer studies. *Journal of Experimental Education*, 51(4), 194-199.

---

## AUTHORS

Robert M. Bernard is an Associate Professor in the Department of Education (Educational Technology), Concordia University, Montreal, Quebec H3G 1M8.

Som Naidu is a doctoral student (Educational Technology) at Concordia University.

ACKNOWLEDGEMENT: The authors wish to thank Dr. Phil C. Abrami and two anonymous reviewers for their careful reading and thoughtful comments.