

ПРАКТИКА МЕТА-АНАЛИЗА И «ВЕЛИКИЕ ДЕБАТЫ О ПЕДАГОГИЧЕСКИХ ТЕХНОЛОГИЯХ»

**Е. Бороховский,
Р.М. Бернارد,
Р.Ф. Шмид
Р.М. Тамим,
Ф.К. Абрами
К. А. Ваде,
М. А. Суркес**

*Университет
Конкордия,
Монреаль,
Канада*

*e-mail:
eborokhovski@education.concordia.ca,
bernard@education.concordia.ca,
schmid@education.concordia.ca,
rana.tamim@education.concordia.ca,
abrami@education.concordia.ca,
anne.wade@education.concordia.ca,
surkes@education.concordia.ca*

Публикация преследует две основные цели:

(1) познакомить читателя с методологией мета-анализа и способствовать популяризации и более широкому использованию мета-аналитических исследований как более информативной и надежной альтернативы традиционным (несистематическим) формам обзоров научно-исследовательской литературы;

(2) представить предварительные данные мета-анализа, изучающего влияние компьютерных технологий на академическую успеваемость и восприятие учебного процесса студентами высших учебных заведений.

Для систематического исследования была использована смешанная модель анализа разнородности данных. Результаты свидетельствуют о следующем:

- использование в учебном процессе компьютерных технологий имеет оптимальный порог, за пределами которого дополнительное насыщение ведет к снижению позитивного влияния на академическую успеваемость;

- наиболее эффективно компьютерные технологии влияют на успеваемость, когда основная цель их использования – поддержка когнитивных процессов а не просто модернизация представления учебного материала;

- сходные по средним значениям результаты объективной оценки успеваемости и субъективной удовлетворенности студентов учебным процессом, проявляют неоднородную динамику под воздействием проанализированных модерлирующих переменных.

Ключевые слова: мета-анализ, компьютерные технологии, высшее образование, величина эффекта, достижение

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 extended from televised instruction on through to applications of computer-based learning, then in its infancy.

Clark's characterization of television as a neutral medium did not come as a particular shock to most media professionals of that era, because of the results of experiments in the 1950s and 1960s where no significant differences in closed-circuit 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 mounting research evidence of the effectiveness of computers in the classroom. A flurry of comments and counter-comments appeared in the literature (e.g., Petkovitch & Tennyson, 1984) and at academic conferences. The issue has re-emerged periodically since 1983, especially in a 1994 special issue of *Educational Technology Research & Development* devoted to it. Luminaries like Clark, Robert Kozma and David Jonassen argued on one side or the other, but the issue was never resolved. Even to this day Clark and others (2009) maintain that features of instructional design and teaching strategies account for the effectiveness of any learning setting, and that technology simply adds convenience, access and cost savings to the complex process of educating and being educated.

While the "technology debate" introduced above represents an ongoing area of concern for teachers and trainers, the focus of this article is on the methodology used to answer such sweeping questions. Clark's claim at the time was based in part on conclusions emerging from several meta-analyses. Meta-analysis, the subject of this article, is a set of techniques that was devised to synthesize the results of potentially hundreds of primary studies. The meta-analyses that preceded Clark's initial claim were conducted prior to 1983, assessing the effectiveness of computer-based instruction versus classroom instruction without computers. While these stud-



ies suggested that technology did have a positive effect on learning, Clark argued that these quantitative summaries were fundamentally flawed because a variety of experimental artifacts, among them the novelty effect associated with the treatment itself, had not been controlled for or factored out of the results. His argument was that if there is no reliable evidence of the existence of a treatment effect in favor of technology, then one must accept the null hypothesis of no effectiveness. Notwithstanding fundamental changes in technology since 1983, Clark holds firm on his contention (Clark, Yates, Early & Moulton, 2009). Was Clark correct in his critique then; is he still right? How do we get closer to the “truth”?

The purpose of this article is twofold: 1) to present and discuss practices and procedures for evaluating research questions using systematic review techniques and meta-analysis; and 2) to use an ongoing meta-analysis of the effectiveness of technology treatments in higher education as an example of these practices. We will discuss the strengths and constraints associated with using meta-analyses to address issues such as those related to the “The Great Technology Debate.” This examination will be informed via a description of how we undertake meta-analyses. We shall use the preliminary results from an ambitious meta-analysis, and examine the extent to which this procedure resolves the question, or indeed raises issues and concerns that make the question(s) more convoluted and intractable.

Methods of Research Synthesis

Why Synthesize 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 hypothesis under consideration. Therefore, it has been common practice for researchers to review the literature of all such studies, whenever enough of these 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 term systematic review has been applied to a review of a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant research, and to collect and analyze data from the studies that are included in the review. Statistical methods (meta-analysis) may or may not be used to analyze and summarize the results of the included studies. Therefore, a systematic review may apply either qualitative or quantitative methodologies to summarize the literature. The generally recognized steps in a systematic review are as follow: 1) specify the research question; 2) specify terms and definitions related to the question; 3) establish inclusion/exclusion criteria; 4) exhaustively search the literature; 5) select studies for inclusion; 6) extract relevant information that will be synthesized based on the synthesis methodology chosen; 7) synthesize the information that has been extracted using the chosen methodology; 8) explore moderator variables (related to meta-analysis only); 9) draw conclusions from the synthesis; 10) disseminate the results. Numbers 6 and 7 in the above list are characterized more generally because they depend on the approach to synthesis that is applied to the collection of studies. Later in this article we will return to these stages in greater detail and illustrate them using our ongoing meta-analysis of the effects of technology on learning outcomes in higher education.

Types of Systematic Reviews

Generally speaking, systematic reviews fall into one of two broad categories: narrative-descriptive or quantitative-descriptive. Narrative reviews attempt to summarize studies descriptively and are usually limited in scope by the sheer volume of prose that is necessary to characterize even a medium-size body of research. Most of the reviews of literature at the beginning of research articles and in the literature review sections of theses are of the narrative type.

There are two general types of quantitative reviews: vote-count reviews and meta-analyses. A vote-count review is simply a frequency count of studies reporting positive, negative or no significant findings resulting from a body of comparative studies (e.g., experimental designs). After studies are collected and classified, a verdict is reached by the plurality of votes that exist in a given category. The vote-count technique 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 compared with smaller samples, but are given equal weight in this technique. Vote count analysis also does not take into account the magnitude of differences/relationships or the quality of the studies that are included in the synthesis.

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 vote-count methodologies that use statistical indices to reflect differential treatment effects (e.g., *t*-values) or relationships among variables (e.g., correlation coefficients). All of the characteristics of a systematic review apply to meta-analysis (i.e., meta-analysis is a type of systematic review), but the methodology of synthesis (i.e., the statistical aggregation of study results) makes it very different from narrative and vote-count reviews.

This article is about meta-analysis, generally speaking, but it uses our review of technology integration in higher education as a case in point. We will describe each phase of a meta-analysis (as a special case of systematic review) and in turn discuss how we accomplished each phase. You will find the headings *General* and *Example* indicating the separation between these sections.

The backdrop of this article, “The Great Technology Debate,” which has raged for years in the literature of educational technology and instructional design, will not be resolved here. In fact, we do not believe that it can be fully resolved empirically. What we hope to accomplish, therefore, is more resolution than convolution—an exposure and discussion of some of the issues that have dogged this debate, and a discussion of some findings related to the application of “modern computer technology” within the structure of the modern academy.

The Practice of Meta-Analysis: Effects of Technology on Learning in Higher Education

The meta-analysis that is described here as an example is part of an ongoing project of the Systematic Review Team at the Centre for the Study of Learning and Performance at Concordia University in Montreal. It was previously published as a research report in *Journal of Computing in Higher Education* (Schmid, et al., 2009). Page numbers for the examples, when exact quotations were used, are from this publication.

Since 1990, more than 30 meta-analyses have been published (Tamim, 2009), each intending to capture the difference between technology-enhanced classrooms and “traditional classrooms” that contain no technology. Many are specific to particular grade levels and subject matters, and some deal with specific forms of technology (e.g., computer-assisted instruction), yet all but three ask the starkly worded question: *Is some technology better than no technology?* The exceptions to this (i.e., Azevedo & Bernard, 1985; Lou, Abrami & D’Apollonia, 2001; Rosen & Salomon, 2007) have attempted to discern among different instructional strategies within technology use. In the modern academy (since about 1990) few classrooms contain no technology, so it makes little sense to continue to cast the question in all-or-none terms. Therefore, the meta-analysis described here takes the unusual approach of attempting to capture the difference between more technology use and less technology use. Since in the majority of the included studies both conditions, treatment and control, contain some technology, it is difficult to apply Clark’s complaint about confounding due to novelty effects to these studies. We also tried to classify technology’s use for teaching and learning purposes to see if there is a difference between, for instance, technologies used as cognitive tools and technologies used to present information. As we constantly remind the reader, this study is incomplete, so the results presented here may change somewhat. But this meta-analysis is a good example for demonstrating how meta-analyses are conducted and can be used by researchers, practitioners and policy makers.

The first nine steps in conducting a systematic review are neither mutually exclusive nor distinct. Rather, they should be viewed as key stages forming part of a continuous and iterative process, each gradually blending into the next stage. The entire systematic review process can take anywhere from six months to several years, depending on the nature, complexity, and scope of the research question(s) being addressed.

1) Specify the research question(s)



General

As in any research effort, research question(s) help to focus attention on the important variables that will be addressed and their relationship to one another. Sometimes this step is quite straightforward and sometimes it can be quite complex. It is appropriate here to search for and examine previous reviews of all kinds. Complexity is introduced when there are synonymous terms and/or fuzzy descriptions of the treatments and dependent measures. For instance, in a recent study (Abrami, et al., 2008) we spent several months examining the relationships between the terms “critical thinking,” “creative thinking,” “problem-solving,” “higher-order thinking” and the like. Likewise, we searched the literature for the range of standardized measures that were available and their psychometric properties. At one point, we conducted a small factor analytical study (Bernard, et al., 2008) of the subscales on one popular standardized measure (i.e., Watson–Glaser Critical Thinking Appraisal; Watson & Glaser, 1980) to determine if they were distinguishable from one another, or whether the test should be considered as a global measure. These are the kinds of issues that sometimes must be dealt with and resolved before the main work of meta-analysis can begin.

Example

- The following research questions were derived from the literature:
- What is the impact of the educational use of contemporary computer-based technologies on achievement outcomes of higher education students in formal educational settings?
- How do various pedagogical factors, especially the amount of technology use and/or the purpose for technology use, moderate this effect?

In relation to the ‘purpose’ cited in the point above, do different uses of technology result in different learning outcomes? (Schmid, et al., 2009, p. 97).

2) Specify terms and definitions related to the question

General

This step continues from Step 1 and involves establishing working or operational definitions of terms and concepts related to the purposes of the meta-analysis. This is done to help further clarify the research questions and to inform the process of devising information search strategies. Definitions also convey what the researchers mean by particular terms, especially when the terms have multiple definitions in the literature. This was the case in the critical thinking project just alluded to in Step 1. This process is critical because a well-defined and clearly articulated review question will have an impact on subsequent steps in the process.

Example

The key terms that frame the research questions above are defined as follows:

- *Educational use* is any use of technology for teaching and learning as opposed to technology serving administrative and managerial purposes.
- *Contemporary computer-based technologies* are more current technologies addressed in studies published since 1990, reflecting the qualitative shift in formal education towards more universal access to the Internet and use of information relevant to establishing and achieving educational goals.
- *Pedagogical factors* refer to the primary functional orientation of technology use. These include 1) promoting communication/interaction and speeding up exchanges of information; 2) providing cognitive support; 3) facilitating information search and retrieval and providing additional sources of information; 4) enriching and/or enhancing the quality of content presentation; and 5) increasing cost-efficiency ratio or otherwise optimizing educational process.
- *Formal educational settings* include instructional interventions of any duration for classroom instruction (CI) in accredited institutions of higher education.
- *Achievement* includes all measures of academic performance. (Schmid, et al., 2009, p. 97).

3) Establish inclusion/exclusion criteria

General

Inclusion/exclusion criteria represent a set of rules that are used both by information specialists to tailor the literature searches, and by reviewers to choose which studies to retain in (or exclude from) the meta-analysis. They also determine the inclusivity or exclusivity of the meta-analysis as a whole. For instance, if the researchers have decided to include only studies of the highest methodological quality (e.g., randomized control trials only), the inclusion criteria will specify this. Likewise, if the review is to have a particular beginning date (e.g., 1990) or include only particular contents or populations, the inclusion/exclusion criteria will indicate this.

Example

This review exercised a liberal approach to study inclusion. Instead of excluding studies of questionable methodological quality, our approach (Abrami & Bernard, 2009) involves coding for research design and other methodological study features to enable subsequent analysis, either by statistically accounting for their differential influence on outcomes or by using weighted multiple regression to remove variance associated with them.

Review for selecting studies for the meta-analysis was conducted in two stages. First, studies identified through literature searches were screened at the abstract level. Then, the review of full-text documents identified at the first stage led to decisions about whether or not to retain each individual study for further analyses.

To be included a study had to have the following characteristics:

- Be published no earlier than 1990.
- Be publicly available (or archived).
- Address the impact of computer technology (including CBI, CMC, CAI, simulations, e-learning) on students' achievements (academic performance) and attitudes (satisfaction with different components of learning experience).
- Be conducted in formal post-secondary educational settings (i.e., a course or a program unit leading to a certificate, diploma, or degree).
- Represent CI [classroom instruction], blended, or computer lab-based experiments, but not distance education environments.
- Contain sufficient statistical information for effect size extraction.

Failure to meet any of these criteria led to exclusion of the study with the reason for rejection documented for further summary reporting. Two researchers working independently rated studies on a scale from 1 (definite exclusion) to 5 (definite inclusion), discussed all disagreements until they were resolved, and independently documented initial agreement rates expressed both as Cohen's Kappa (κ) and as Pearson's r between two sets of ratings. (Schmid, et al., 2009, p. 99).

4) Search the literature for studies that contain the specified terms

General

This is arguably one of the most important aspects of conducting a systematic review/meta-analysis, as it may be compared to the data collection phase of a primary study. To meet the criterion of comprehensiveness and minimize what is known as the "publication bias" phenomenon, it is necessary to look beyond the published literature to the "grey literature" found in conference presentations, dissertations, theses, reports of research to granting agencies, government agencies, etc, and in the "file drawers" of researchers who may possess unpublished manuscripts. A diversity of bibliographic and full-text databases must be searched, including those in related fields and geographic regions. Since different fields (and cultures) use different terminology, strategies for each database must be individually constructed. In addition to the database searches, web searches for grey literature, manual searches through the tables of contents of the most pertinent journals and conference proceedings, and branching from previous review articles or selected manuscripts should also be conducted. In some cases researchers will contact prominent and knowledgeable individuals in the field to determine if they know of additional works that fit the inclusion/exclusion criteria. Literature searches may continue even as other stages in the review are proceeding, so that the process of information search and re-



trieval is best described as iterative. Naive information retrieval will result in a systematic review that has limited generalizability or, even worse, biased results, hence use of the services of a professional information specialist is strongly advised.

Example

Extensive literature searches were designed to identify and retrieve primary empirical studies relevant to the major research question. Key terms used in search strategies, with some variations (to account for specific retrieval sources), primarily included: “technolog*”, “comput*”, “web-based instruction,” “online,” “Internet,” “blended learning,” “hybrid course*,” “simulation,” “electronic,” “multimedia” OR “PDAs” etc.) AND (“college*,” “university,” “higher education,” “postsecondary,” “continuing education,” OR “adult learn*”) AND (“learn,*” “achievement*,” “attitude*,” “satisfaction,” “perception*,” OR “motivation,” etc.), but excluding “distance education” or “distance learning” in the subject field. To review the original search strategies, please visit <http://doe.concordia.ca/cslp/>.

The following electronic databases were among those sources examined: ERIC (Web-Spurs), ABI InformGlobal (ProQuest), Academic Search Premier (EBSCO), CBCA Education (ProQuest), Communication Abstracts (CSA), EdLib, Education Abstracts (WilsonLine), Education: A SAGE Full-text Collection, Francis (CSA), Medline (PubMed), ProQuest Dissertation & Theses, PsycINFO (EBSCO), Australian Policy Online, British Education Index, and Social Science Information Gateway.

In addition, a Google Web search was performed for grey literature, including a search for conference proceedings. Review articles and previous meta-analyses were used for branching, as well as the table of contents of major journals in the field of technology (e.g., *Educational Technology Research & Development*). (Schmid, et al., 2009, pp. 98-99)

5) Select studies for inclusion

General

In this step raters apply the inclusion/exclusion criteria to the studies that have been retained through searches. The first step normally involves an examination of abstracts, so as to avoid the cost of retrieving full text articles on the first step. The next step is to retrieve full text documents for further examination. Again, raters apply the inclusion/ exclusion criteria as they examine the entire document for relevance. Normally, two raters are used to accomplish these selection tasks and inter-rater reliability is calculated to indicate the degree of agreement between them.

Example

This is a meta-analysis in progress; hence we call it a Stage I review. Overall, more than 6,000 abstracts were identified and reviewed, resulting in full-text retrieval of about 1,775 (close to 30% of the original dataset) primary research studies potentially suitable for the analysis. Out of this number, through a thorough review of full-text documents, 490 studies were retained for further analysis. To date, they yielded 681 effect sizes, 541 of which were in the *Achievement* category. We anticipate that these numbers represent approximately up to half of the studies that will eventually be processed.

Eight outliers with effect sizes exceeding ± 2.5 SD were excluded from the distribution (five in the *Achievement* category). Also, a sample size of 36,943 students in the control condition was lowered to match the number of students in the experimental condition ($N = 1,156$) in order to reduce the leverage that this study would have exerted on the weighted averages.

Subsequently, 536 ($N = 41,159$) achievement effect sizes were retained for the overall analysis. Also, in this article we report results of the moderator variable analyses of the effects for which study features coding was completed. There were 310 achievement effect sizes (fully coded for study features), based on a total sample of 25,497 participants in treatment and control conditions

Inter-rater agreements at different stages of the review were as follows:

- Screening abstracts—87.69% (Cohen’s $\kappa = 0.75$) or $r = 0.77, p < .001$.
- Full-text manuscript inclusion/exclusion—84.58% ($\kappa = 0.69$) or $r = 0.81, p < .001$.

- Attributing dimensions and estimating magnitude of the difference between conditions—75.44% ($\kappa = 0.51$) based on sample coding of 100 studies at the beginning of the project.
- Effect size extraction—91.90% ($\kappa = 0.84$).
- Study features coding—91.25% ($\kappa = 0.83$). (Schmid, et al., 2009, p. 98).

6) Extract effect sizes and study feature information

General—Determining the Direction of the Effect

Before effect sizes can be extracted the researcher must determine which condition will be designated as the treatment group and which will be designated as the control group. In most meta-analyses, designation of the treatment or intervention group and the control group is clear. Usually, the treatment group receives the intervention and the control group does not. A good example of this is the meta-analysis by Bernard, et al. (2004), in which distance education conditions (the treatment) were compared to classroom instruction conditions (the control). There are some circumstances, especially when two treatments are being compared, when this designation is not clear. The question being asked is: *Which of these two treatments is most effective?* In this circumstance, it is necessary to have some framework or rational basis for establishing what characteristics of the intervention will define the “treatment” and what characteristics will define the “control.” In Bernard, et al. (2009) different interaction treatments in distance education environments were compared. The intervention condition was determined to be the condition that likely evoked the most interaction between 1) students and other students, 2) students and the teacher; and 3) students and the content to be learned. We encountered a similar problem in this study and the following describes how we dealt with it.

Example

The degree of technology saturation was used to determine the distinction between the experimental and control conditions. Conditions that were more highly saturated were designated as the experimental conditions while the less saturated conditions were considered the control.

Technology saturation was defined as follows:

- Greater intensity of use (more frequently and/or for longer time periods)
- More advanced technology (whether it features more/fewer options, or enables more/fewer different functions)

Use of more types of tools, devices, software programs, etc., so that the summative exposure to various technologies is higher in the experimental condition than it is in the control condition.

Experimental and control group designations based on these rules were conducted by independent coders using procedures similar to those described above. (Schmid, et al., 2009, p. 100).

The degree of difference in technology saturation between experimental and control conditions was coded on a 3-point scale: minimal, moderate, and high.

General—Effect size Extraction and Calculation

To avoid the problems associated with inferential statistics (see reference to Vote Count methodology), Gene Glass developed a standardized metric, referred to as an effect size, for quantitatively describing the difference between a treatment or intervention condition and a control condition. It involves subtracting the control group mean from the treatment group mean, and then standardizing this difference by dividing it by the standard deviation of the control group. A positive valence on an effect size indicates that the treatment group has outperformed the control group. A negative valence means the reverse. Equation 1 shows the form of this metric, called Glass' Δ .

Effect size extraction is defined as the process of locating and coding information contained in research reports that allows for the calculation of an effect size. In the best instance, this information is in the form of a mean, standard deviation and sample size for the experimental and the control conditions. There is also a modification of the basic effect size equation for studies reporting pretest and posttest data for both experimental and control groups (Boren-



stein, Hedges, Higgins & Rothstein, 2009). In other cases, effect sizes can be estimated from inferential statistics (e.g., t-tests, F-tests, or p-levels) using conversion formulas provided by Glass, McGaw and Smith (1981) and Hedges, Shymansky and Woodworth (1989).

$$\Delta = \frac{\bar{X}_E - \bar{X}_C}{SD_C} \tag{Equation 1}$$

Cohen (1988) modified this equation to represent the joint variation in the treatment and the control groups by producing an effect size metric (called Cohen’s *d*) based on division of the mean difference by the pooled standard deviations of both groups (Equation 2). Cohen also provides a rough guide to interpreting an effect size. Effect sizes up to about 0.20 *SD* represent a small advantage for the treatment group over the control group, whereas an effect size of around 0.50 is a medium advantage. Effect sizes of 0.80 and over are considered to be a large advantage.

$$d = \frac{\bar{X}_E - \bar{X}_C}{SD_{Pooled}} \tag{Equation 2}$$

The pooled or within-group standard deviation (*SD_{pooled}*) is calculated as follows.

$$SD_{Pooled} = \sqrt{\frac{(n_E - 1)SD_E^2 + (n_C - 1)SD_C^2}{n_E + n_C - 1}} \tag{Equation 3}$$

Based on an analysis of studies of effect sizes with different sample sizes, Hedges and Olkin (1985) concluded that studies with small samples tended to overestimate the true effect size, and to correct for this, developed an adaptation of Cohen’s *d* that contains a correction for bias when *N* is less than about 40. This corrected effect size, called Hedges’ *g*, is shown in Equation 4. Degrees of freedom (*df*) is the same denominator term used in calculating the pooled standard deviation in Equation 3.

$$g = d \left(1 - \frac{3}{4df - 1} \right) \tag{Equation 4}$$

The standard error (*SE*) of *g* (the standard deviation of the sampling distribution of *g*) is then found using Equation 5. Larger samples will produce standard errors that are smaller than in smaller samples, reflecting the fact that larger samples contain less measurement error than smaller samples.

$$SE_g = \sqrt{\left(\frac{n_E + n_C}{n_E n_C} \right) + \left(\frac{d^2}{2(n_E + n_C)} \right) \times \left(1 - \frac{3}{4df - 1} \right)} \tag{Equation 5}$$

The variance for *g* is found by squaring *SE_g*. This is shown in Equation 6.

$$V_g = SE_g^2 \tag{Equation 6}$$

Using *g* and the standard error of *g*, the upper and lower boundaries of the 95th confidence interval (*CI*) can be constructed. Larger samples will produce a smaller interval and smaller samples will produce a larger interval. Equation 7 shows how the upper and lower boundaries of this interval are calculated. If this confidence interval does not contain the null value (0.0), then the effect size is interpreted as exceeding what would be expected by chance when $\alpha = .05$.

$$\begin{aligned} Lower &= g - 1.96 \times SE_g \\ Upper &= g + 1.96 \times SE_g \end{aligned} \tag{Equation 7}$$



Another way of testing the null hypothesis is to calculate a Z -test. This is accomplished by dividing the effect size g by the standard error, as shown in the following equation.

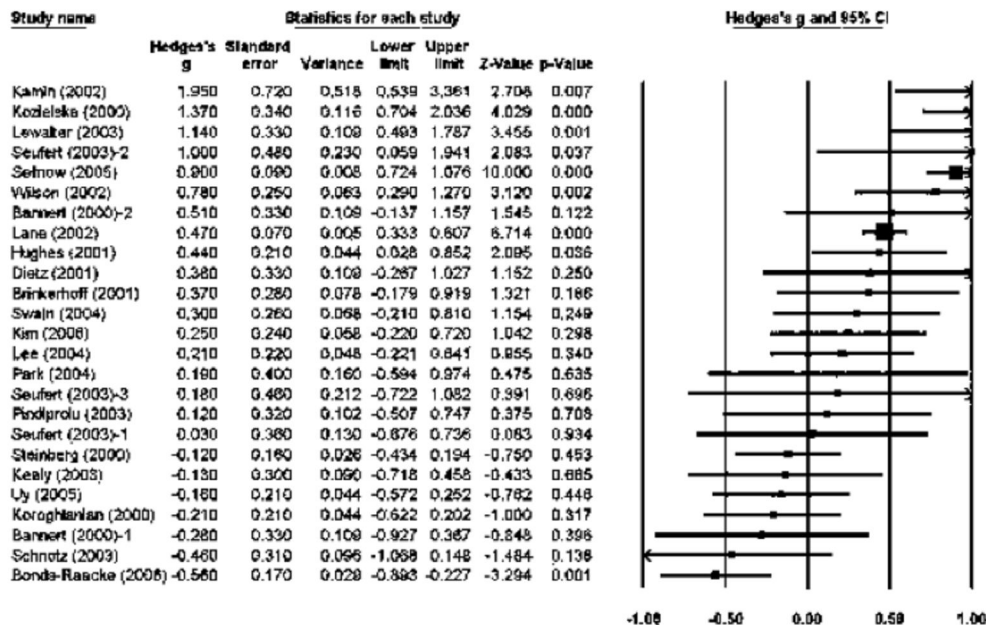
$$Z_g = \frac{g}{SE_g} \quad (\text{Equation 8})$$

This value is then compared to $Z = 1.96$, the standard deviation describing the upper boundary of the 95% region on the unit normal distribution.

Example

As a demonstration of how the basic statistics just presented appear in the software package *Comprehensive Meta-Analysis* (Borenstein, Hedges, Higgins & Rothstein, 2005), Figure 1 shows the descriptive statistics associated with a subset of 25 effect sizes drawn from the complete distribution of 536 effect sizes. On the right side of the figure is a graphic representation called a Forest plot. The effect size for each study is depicted as a dot. The lines around it show the width of the 95th confidence interval. Note that confidence intervals spanning 0.0 on the distribution are considered to be not significantly different from zero. The Z -test of these effect sizes also indicates that $p > .05$. The dots that represent the effect size vary in size. Smaller dots are lower leverage effect sizes (i.e., smaller contributors to the average effect size), while larger dots are higher leverage effects characterized by larger sample sizes.

Figure 1. Forest plot of 25 effect sizes from the complete distribution of effect sizes.



General—Study Feature Coding

Study features are characteristics of studies that range across a wide number of studies in the total distribution. Generally speaking, there are three groups of study features that can be considered: 1) publication characteristics (e.g., date of publication, type of publication); 2) methodological characteristics (e.g., research design, measurement validity); and 3) substantive characteristics (e.g., type of intervention, age of participants).

Example

Moderator analysis of coded study features was used to explore variability in effect sizes. These study features derived from an ongoing analysis of the theoretical and empirical literature in the field and were based on our previous meta-analyses (Bernard, Abrami, Borokhovski et al., 2009). Study features were of three major categories: methodological (e.g., research design), demographic (e.g., type of publication), and substantive (e.g., purpose of technology use).



Among the latter, we were most interested in the study feature describing differential functionality (or major purposes) of technology use. We considered the following dimensions on which experimental conditions consistently could be contrasted to control conditions: Immediacy of Communication, Cognitive Support, Provision of Additional Sources of Information (including Information Search and Retrieval Tools), Presentation/Delivery Enhancement, and Optimization of Educational Processes where technology was not directly involved in the learning process, but addressed cost-efficiency and other aspects of administration of education. (Schmid, et al., 2009, p. 101)

7) Synthesize the effect sizes that have been extracted

General

After all of the effect sizes have been extracted and the basic statistics calculated, the next step is to synthesize them. This involves finding the average effect size ($g+$) and accompanying statistics. Normally, the statistics of most interest are the same as those calculated for each effect size—the standard error, the variance, the upper and lower limits of the 95th confidence interval and the Z -value. In addition, a Q -value is also calculated and tested using the χ^2 -distribution with $p - 1$ degrees of freedom.

In calculating the average effect size under the fixed effect model, the first step is to calculate a weight (W) for each study that reflects the size of the sample. Under the fixed effect model, larger studies are given more weight than smaller samples. The weight is calculated using the inverse variance method shown in Equation 9. The variance (V_i) and the resulting weights (W_i) are shown in Table 1. Note that the Sellnow (2005) and Lane (2002) studies are given disproportionately greater weight than the other studies. By contrast, under the random model of effect size synthesis, studies of different sizes are given more equal representation in the analysis. The individual weights are then summed (ΣW_i).

$$W_i = \frac{1}{V_i} \tag{Equation 9}$$

The next value needed is the product of g_i times W_i . These individual values are shown in the last column of Table 1. These values are summed ($\Sigma g_i W_i$). The average effect size for the distribution is then calculated by dividing $\Sigma g_i W_i$ by ΣW_i as in Equation 10.

$$g+ = \frac{\sum g_i W_i}{\sum W_i} \tag{Equation 10}$$

For this sample set of 25 effect sizes, the average effect size is as follows.

$$g+ = \frac{248.10}{652.43} = 0.38$$

Table 1. Effect size statistics needed to calculate summary statistics

Study name	g_i	V_i	W_i	$(g_i)(W_i)$
Kamin (2002)	1.95	0.52	1.93	3.76
Kozielska (2000)	1.37	0.12	8.65	11.85
Lewalter (2003)	1.14	0.11	9.18	10.47
Seufert (2003)-2	1.00	0.23	4.34	4.34
Sellnow (2005)	0.90	0.01	123.46	111.11
Wilson (2002)	0.78	0.06	16.00	12.48
Bannert (2000)-2	0.51	0.11	9.18	4.68
Lane (2002)	0.47	0.00	204.08	95.92

Hughes (2001)	0.44	0.04	22.68	9.98
Dietz (2001)	0.38	0.11	9.18	3.49
Brinkerhoff (2001)	0.37	0.08	12.76	4.72
Swain (2004)	0.30	0.07	14.79	4.44
Kim (2006)	0.25	0.06	17.36	4.34
Lee (2004)	0.21	0.05	20.66	4.34
Park (2004)	0.19	0.16	6.25	1.19
Seufert (2003)-3	0.18	0.21	4.73	0.85
Pindiprolu (2003)	0.12	0.10	9.77	1.17
Seufert (2003)-1	0.03	0.13	7.72	0.23
Steinberg (2000)	-0.12	0.03	39.06	-4.69
Kealy (2003)	-0.13	0.09	11.11	-1.44
Uy (2005)	-0.16	0.04	22.68	-3.63
Koroghlanian (2000)	-0.21	0.04	22.68	-4.76
Bannert (2000)-1	-0.28	0.11	9.18	-2.57
Schnotz (2003)	-0.46	0.10	10.41	-4.79
Bonds-Raacke (2006)	-0.56	0.03	34.60	-19.38
Sums			652.43	248.10

Example

Data were analyzed using *Comprehensive Meta-Analysis* (Borenstein, et al., 2005), a dedicated meta-analysis software. The summary statistics derived from 536 effect sizes are shown in Table 2. In addition to the $g+$ of 0.25, the lower and upper limits of the 95th confidence interval of the Z -distribution are shown along with the variability statistics Q (Total) and I -squared. (For the equations and more information, consult Hedges & Olkin, 1985; Borenstein, et al., 2009.) The Q -statistic tells us that the distribution is significantly heterogeneous ($p < .001$, and the I -squared statistic indicates that 73.20% of the variability in effect sizes exceeds sampling error (i.e., variability expected from chance fluctuation due to sampling).

The significant variability suggests that differences among samples exist, and that some of this fluctuation may be accounted for by identifiable characteristics of studies. Under this condition of heterogeneity, moderator analysis is warranted.

Table 2. Overall weighted average effect size and homogeneity statistics

Outcome	k	$g+$	Lower 95th	Upper 95th	Q (Total)	I -squared
Achievement	536	0.25*	0.23	0.27	1,996.40**	73.20

* $p < .05$, ** $p < .01$ (Schmid, et al., 2009, p. 102).

8) Explore moderator variables (if appropriate)

General

Moderator variable analysis can take two forms: 1) analysis of categorical variables using a meta-analysis analog to analysis of variance, and 2) weighted meta-regression of continuous predictor variables. In the former case, the researcher might be interested in exploring the between-group effects of research design (i.e., a coded categorical variable) to determine if different designs produce different results. In the latter case, the question might arise as to the relationship of "year of publication" and effect size. Both statistical procedures apply the weights that were generated in the last section, so that larger studies get more weight (or importance) than smaller studies.



Example—Categorical Moderator Analysis

Moderator variable analysis of coded study features attempts to identify the systematic sources of variation. One potential source of variation that is often referred to in the literature of meta-analysis (e.g., Abrami & Bernard, 2006) derives from the presence of different research designs. True experiments employing random assignment intended to neutralize selection bias are the most highly prized. Quasi-experimental designs that employ pretesting to establish group equivalence are considered to be reasonable alternatives to true experiments. Pre-experimental designs that contain no mechanism to ensure that selection bias is controlled are considered to be the weakest form of evidence (Campbell & Stanley, 1963).

We are of the view that all forms of evidence are acceptable, as long as their equivalence can be demonstrated. If it cannot, Abrami and Bernard (2009) have proposed a methodology for equalizing studies of different methodological quality while leaving within-group variability intact. The decision to adjust categories of studies is based on a between-group test of these studies. As an initial step, the three categories of research design identified in the coding of study features were compared. The top third of Table 3 shows the results of this mixed effect analysis. The three categories of research design were not significantly different ($p = .47$), so we decided not to employ the adjustment procedure alluded to earlier.

Table 3. Mixed analysis of three moderator variables: Research Design, Level of Technology Saturation, and Purpose of Technology Use.

Moderator Variables (Mixed model)	<i>k</i>	<i>g</i> ⁺	Lower 95th	Upper 95th	<i>Q_B</i>
Research Design					
Pre-Experiments	82	0.32	0.23	0.41	
Quasi-Experiments	56	0.32	0.21	0.44	
True Experiments	172	0.25	0.16	0.33	
Between Classes					1.52
Level of Technology Saturation (Strength)					
Low Saturation	143	0.33	0.29	0.37	
Medium Saturation	116	0.29	0.25	0.33	
High Saturation	51	0.14	0.08	0.20	
Between Classes					8.59*
Type Purpose of Technology Use					
Cognitive Support	112	0.41	0.36	0.45	
Presentational Support	89	0.10	0.05	0.15	
Multiple Uses	78	0.29	0.24	0.34	
Between Classes					17.05*
Total (Fixed effect)	310	0.28	0.25	0.30	1,131.54

* $p < .05$

The results of the analysis of two other study features are shown in the lower two-thirds of Table 3. Level of technology saturation, defined previously in the Method section, produced a significant between-group effect. When these levels were examined using Bonferroni post hoc tests adapted for meta-analysis (Hedges & Olkin, 1985), it was found that low and medium saturation formed a homogeneous set that was significantly different from the category of studies exhibiting high saturation. The combined average of low and medium saturation ($k = 159$) is

0.32. This result suggests that more technology use is not necessarily better in encouraging achievement in higher education classrooms. (Schmid, et al., 2009, pp. 102-103)

Example—Meta-Regression Analysis

In the final analysis, we wanted to know if the effects of technology on achievement had changed during the period covered by this meta-analysis. To do this we ran weighted multiple regression, treating Publication Year as the predictor and Hedges' g as the outcome variable. The results of this analysis revealed that effects associated with technology use in higher education ($k = 310$) have not changed substantially over the years ($\beta = 0.0034$, $p = .28$, $Q_R = 1.88$). This relationship is displayed graphically in Figure 1. There is wide variability around the average effect of 0.28, but the regression line is virtually flat. We would like to reiterate that these results are preliminary. When more effect sizes are added, and if this trend continues, we may conclude that in spite of advances in computer hardware, software, and especially the presence of the Internet, the impact of such tools on learning achievement has not dramatically changed over the period covered by this meta-analysis. (Schmid, et al., 2009, pp. 103-104).

Figure 2. Scatterplot of publication year by effect size ($k = 310$).

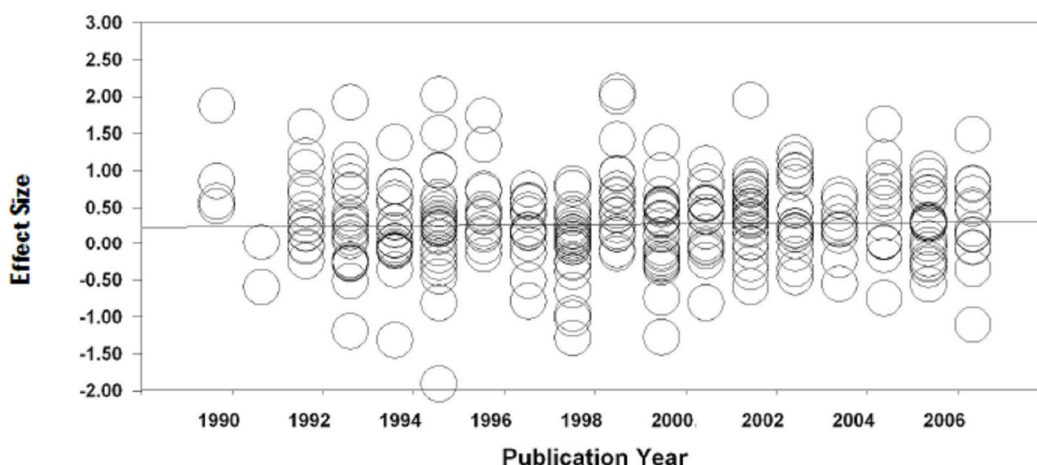


Figure used with the permission of Springer Publishing Company.

9) Draw conclusions from the meta-analysis

General

This step in a meta-analysis is fairly straightforward as it establishes how collected and analyzed data inform the research question(s) that guided the review and explores the possible conceptual and practical implications.

Example

The most general conclusion that can be drawn from this partially completed meta-analysis is that the effect of technology on learning is modest, at best, in higher education classrooms where some technology is present in both treatment and control conditions. The average effect size of 0.25 ($k = 536$) is somewhat lower than the average effect that Tamim (2009) found ($g+ = 0.32$) where 25 meta-analyses of technology's effect on learning (dated from 1996 to 2008) were combined in a second order meta-analysis. This study included 1,055 primary studies involving approximately 109,700 students. This difference is somewhat predictable given the "all vs. none" approach of previous meta-analyses compared with the "something vs. something" approach of this meta-analysis. The heterogeneity results of this meta-analysis indicated a broad range in effect sizes ($Q = 1,996.40$, $p < .001$) with 73.20% (I^2) of the heterogeneity deriving from differences among studies (i.e., that exceeding sampling error). This brings into question the exact location of the point estimate.



Analysis of four categorical moderator variables, one methodological and two substantive, suggested the following results: 1) there is no difference between the three classes of research designs included in the collection (i.e., pre-experimental, quasi-experimental and true experimental designs); 2) technology saturation levels differed, with low and medium being about equal (i.e., $g+ = 0.37$ and 0.33 , respectively), but significantly different from high saturation (i.e., $g+ = 0.14$); and 3) cognitive support uses of technology ($g+ = 0.41$) significantly outperformed presentational uses ($g+ = 0.10$). Meta-regression of “date of publication” on effect size resulted in an almost flat regression line that was not significantly different from zero.

All of the substantive results bear on questions related to “the great technology debate” alluded to earlier in this article. First, if there is the expectation of the presence of novelty effects (i.e., those receiving technology treatments performing better because it is novel compared to normal classroom activities), it is reduced by the fact that neither treatment nor control condition contained “no technology.” Second, the “novelty effect explanation” would likely predict increasing effects based on increasing novelty. This was found not to be the case. Lower and medium uses of technology produced larger effects than highly saturated classroom uses. Third, Clark’s hypothesis concerning technology’s neutral effect on learning would predict that all uses of technology would be about equal, with no advantage for cognitive support tools over presentational tools. This was found not to be the case, with cognitive tool uses outperforming presentational tool uses (i.e., $g+ = 0.41$ vs. 0.10 , respectively). These results, by no means resolve “the great technology debate,” but they do add some measure of clarity to Clark’s stark assessment of the effects of technology role in the higher education classroom.

The results of the regression analysis present a rather depressing note to the effects of the classroom use of technology over the years 1990 to the present. In spite of the rapid growth of the Internet, the greater sophistication of computer tools and the role of technology in enhanced communication, there is no evidence that the application of these advances translates into better achievement outcomes. This suggests that technology has not transformed teaching and learning in the ways that Robert Kozma (1994) and others predicted. That, of course, does not mean that it won’t. The effects of any innovation may lag behind widespread adoption by the general public, and even more specifically, the widespread effect on some indirect measure of effect like student achievement.

Though attitude data are generally beyond the scope of this article, it is probably worth mentioning that, contrary to the achievement outcomes, results in the attitude category showed slightly different dynamics. For example, among three levels of technology saturation, the highest effect on attitudes was associated with the middle category, which could be suggestive of something similar to what is known as the Yerkes-Dodson (1908) law in psychophysiology applied to the domain of student motivation. Too little exposure to modern technology fails to induce interest in learners, while too much overwhelms them to the point of distraction from learning objectives and even possible dissatisfaction with the learning process. Somewhat related to the above, is another issue of high consequence that continuously attracts educators’ attention: to what extent satisfaction or dissatisfaction with instruction, instructor, and instructional circumstances and qualities helps or impedes student learning, or, in other words, how “liking” and “learning” are related in formal education. Attempts to address this issue are plentiful. As far back as the 1970s, an experimental study by Clifford (1973) reasonably attributed positive correlations between achievement and attitude outcomes to students’ task commitment. Data supporting (e.g., Knuver & Brandsma, 1993) or questioning (e.g., Delucchi & Pelowski, 2000) positive relationships between affective and cognitive outcomes in education have been published since, though they understandably focused more on aspects of student motivation. What other salient factors (associated with instructional design and delivery) and in what combinations may be responsible for building reciprocity between enjoying learning process and achieving learning objectives?

Recently, in the domain of distance education, Anderson (2003) formulated a hypothesis according to which student will benefit from both “deep and meaningful” learning and “more satisfying educational experience” when they are given sufficient opportunities for interacting



among themselves, with the teacher and/or course content, a hypothesis that was addressed and largely confirmed in Bernard et al. (2009). Similarly, in-class technology implementation can be guided by some basic principles that make liking and learning work together to mutual benefit, and the best, if not the only way of searching for these principles lies in the methodology of meta-analysis when a large collection of studies that report both types of outcomes is available for a systematic review. Naturally, we intend to further explore the relationships between attitude and achievement outcomes in the entire collection of studies in this meta-analysis.

10) Disseminate the results

There are three audiences that may be interested in the results of a systematic review of the type described here. Practitioners, teachers in higher education contexts in this instance, may use the results to become knowledgeable of research findings and possibly modify their practices. Policy-makers, individuals who make purchase decisions and form policies that affect the large-scale adoption of innovations, may be informed by the results of a systematic review. It is recognized that research represents only one form of evidence for decision-making, but with access to the voices of researchers, broadly characterized, policy-makers are in a better position to make rational and informed choices. The third group that is likely to be affected by a systematic review is researchers who have contributed, or may potentially contribute studies to the growing corpus of evidence that form the substance of reviews. Researchers need to know the directions of inquiry that are informing their field, in addition to the design, methodological and reporting conventions that make it possible for individual studies to be included in reviews.

For the first two groups, there may be an issue related to the form of reporting a systematic review or meta-analysis. As has been demonstrated here, a fair degree of knowledge is required to construct a review, and likewise there is some degree of knowledge related to its interpreting and applying its findings. Knowledge translation, sometimes referred to as *knowledge mobilization*, is often desirable. In many domains (e.g., health), organizations and centers have been established to act as “go-betweens,” of sorts, linking the researcher or meta-analyst with consumers of the results of systematic reviews or meta-analyses (e.g., Cochrane Collaboration, <http://www.cochrane.org>; CSLP Knowledge Link, <http://doe.concordia.ca/cslp/RS-Publications-KL.php/>). In addition, organizations such as the Campbell Collaboration (<http://www.campbellcollaboration.org/>) have been established for the primary purpose of maintaining review standards and acting as a repository for high-quality reviews.

Conclusion

In this article we have provided a brief description of meta-analysis as a methodology for quantitatively synthesizing the results of many comparative studies organized around a central question or set of questions. Ten steps were described in moderate detail and an example, the application of technology to teaching in higher education (Schmid, et al., 2009), was provided. Using this methodology to address the question of whether technology has an impact on learning, we have found numerous examples of how meta-analysis both effectively synthesizes extant, empirical data, and perhaps even more importantly, serves as a heuristic to identify potent, causal factors that can inform practice and further research. We recognize that such procedures are prisoners of the data: if the data are biased, so too will the results be. This illustrates the imperative that research unequivocally isolates variables, measures them with valid and reliable tools, and restricts the interpretation of outcomes to those that are core to the research question. That said, most research falls short on one or more of these criteria. As such, by looking at scores or hundreds of studies, we reduce the errors in conclusions though never eliminating them completely.

More information about specific topics concerning systematic review and meta-analysis is available in the following publications: 1) history (Hunter, 1997; Glass, 2000); 2) general methodology (e.g., Cooper, 1989; Cooper & Hedges, 1994; Lipsey & Wilson, 2001; Pettigrew & Roberts, 2006); and 3) statistics (e.g., Hunter & Schmidt, 1990; Hedges & Olkin, 1995; Borenstein, Hedges, Higgins & Rothstein, 2009).



References

Note: A list of studies included in the meta-analysis is available upon request from the authors.

- Abrami, P.C., Bernard, R.M., Borokhovski, E., Wade, A., Surkes, M., Tamim, R., & Zhang, D. A. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage one meta-analysis. *Review of Educational Research*, 78(4), 1102-1134. doi:10.3102/0034654308326084
- Abrami, P.C., & Bernard, R.M. (2009). *Statistical control vs. classification of study quality in meta-analysis*. Manuscript submitted for publication.
- Anderson, T. (2003). Getting the mix right again: An updated and theoretical rationale for interaction. *International Review of Research in Open and Distance Learning*, 4(2), 9-14. Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/149>
- Azevedo, R., & Bernard, R.M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, 13(2), 111-127. doi:10.2190/9LMD-3U28-3A0G-FTQT
- Bernard, R.M., Abrami, P.C., Borokhovski, E., Wade, C.A., Tamim, R., Surkes, M.A., & Bethel, E.C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*. Advance online publication. doi:10.3102/0034654309333844
- Bernard, R.M., Abrami, P.C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., Waiet, P. A., Fiset, M., & Huang, B. (2004). How does distance education compare to classroom instruction? A Meta-analysis of the empirical literature. *Review of Educational Research*, 74(3), 379-439. doi:10.3102/00346543074003379
- Bernard, R.M., Zhang, D., Abrami, P.C., Sicol, F., Borokhovski, E., & Surkes, M. (2008). Exploring the structure of the Watson-Glaser Critical Thinking Appraisal: One scale or many subscales? *Thinking Skills and Creativity*, 3, 15-22. doi:10.1016/j.tsc.2007.11.001
- Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley & Sons.
- Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H. (2005). *Comprehensive Meta-analysis Version 2*. Englewood, NJ: Biostat.
- Campbell, D., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.
- Clark, R.E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53(4), 445-459. doi:10.3102/00346543053004445
- Clark, R.E. (1994). Media will never influence learning. *Educational Technology Research & Development*, 42(2), 21-29. doi:10.1007/BF02299088
- Clark, R.E., Yates, K., Early, S., & Moulton, K. (2009). An analysis of the failure of electronic media and discovery-based learning: Evidence for the performance benefits of guided training methods. In K. H. Silber, & R. Foshay, (Eds.). *Handbook of training and improving workplace performance, Volume I: Instructional design and training delivery*. Washington, DC: International Society for Performance Improvement.
- Cobb, T. (1997). Cognitive efficiency: Toward a revised theory of media. *Educational Technology Research & Development*, 45(4), 21-35. doi:10.1007/BF02299681
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cooper, H.M. (1989). *Integrating research: a guide for literature reviews* (2nd ed.). Newbury Park, CA: SAGE Publications.
- Cooper, H. & Hedges, L.V. (Eds.) (1994). *The Handbook of Research Synthesis*. New York: Russell Sage.
- Glass, L.V. (2000). *Meta-analysis at 25*. Retrieved on May 21, 2008 from: <http://glass.ed.asu.edu/gene/papers/meta25.html>
- Glass, G.V., McGaw, B., & Smith, M.L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Clifford, M. M. (1973). How learning and liking are related: A clue. *Journal of Educational Psychology*, 64(2), 183-186. doi:10.1037/h0034587
- Delucchi, M., & Pelowski, S. (2000). Liking or learning? The effect of instructor likeability and student perceptions of learning on overall ratings of teaching ability. *Radical Pedagogy*, 2(2). Retrieved from http://radicalpedagogy.icaap.org/content/issue2_2/delpel.html
- Hedges, L.V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hedges, L.V., Shymansky, J.A., & Woodworth, G. (1989). *A practical guide to modern methods of meta-analysis*. (Stock Number PB-52). Washington, DC: National Science Teachers Association. (ERIC Document Reproduction Service No. ED309952).

- Higgins, J.P.T., Thompson, S.G., Deeks, J.J., & Altman, D.G. (2003). Measuring inconsistency in meta-analysis. *British Medical Journal*, 327, 557-560. doi:10.1136/bmj.327.7414.557
- Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. NY: Russell Sage Foundation.
- Hunter, J.E. & Schmidt, F.L. (1990). *Methods of meta-analysis: correcting error and bias in research findings*. Newbury Park, CA: SAGE Publications.
- Jonassen, D., Campbell, J. and Davidson, M., (1994). Learning with media: Restructuring the debate. *Educational Technology Research & Development*, 42(2), 31-39.
- Kozma, R. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research & Development*, 42(2), 7-19. doi:10.1007/BF02299087
- Knuver, A. W. M., & Brandsma, H. P. (1993). Cognitive and affective outcomes in school effectiveness research. *School Effectiveness and School Improvement*, 4(3), 189-204. doi:10.1080/0924345930040302
- Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis*. New York, NY: Sage Publications.
- Lou, Y., Abrami, P. C., & D'Appollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, 71, 449-521. doi:10.3102/00346543071003449
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Oxford, UK: Blackwell Publishing.
- Rosen, Y., & Salomon, G. (2007). The differential learning achievements of constructivist technology-intensive learning environments as compared with traditional ones: A meta-analysis. *Journal of Educational Computing Research*, 36(1), 1-14. doi:10.2190/R8M4-7762-282U-554J
- Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R., Abrami, P. C., Wade, A., Surkes, M. A., & Lowerison, G. (2009). Technology's effect on achievement in higher education: A stage I meta-analysis of classroom applications. *Journal of Computing in Higher Education*. 21(2), 95-109. doi:10.1007/s12528-009-9021-8
- Tamim, R. M. (2009). *Effects of Technology on Students' Achievement: A Second-Order Meta-Analysis* (Unpublished doctoral dissertation). Concordia University, Montreal, QC, Canada.
- Watson, G., & Glaser, E. M. (1980). *Watson-Glaser Critical Thinking Appraisal: Forms A and B*. San Antonio, TX: PsychCorp.
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482. doi:10.1002/cne.920180503

THE PRACTICE OF META ANALYSIS AND «THE GREAT TECHNOLOGY DEBATE»

**E. Borokhovski,
R.M. Bernard,
R.F. Schmid,
R. M. Tamim,
P.C. Abrami,
C.A. Wade,
M. A. Surkes**

Concordia University,
Canada

e-mail:
eborokhovski@education.concordia.ca,
bernard@education.concordia.ca,
schmid@education.concordia.ca,
rana.tamim@education.concordia.ca,
abrami@education.concordia.ca,
anne.wade@education.concordia.ca,
surkes@education.concordia.ca

In this article we have provided a brief description of meta-analysis as a methodology for quantitatively synthesizing the results of many comparative studies organized around a central question or set of questions. Ten steps were described in moderate detail and an example, the application of technology to teaching in higher education (Schmid, et al., 2009), was provided. Using this methodology to address the question of whether technology has an impact on learning, we have found numerous examples of how meta-analysis both effectively synthesizes extant, empirical data, and perhaps even more importantly, serves as a heuristic to identify potent, causal factors that can inform practice and further research. We recognize that such procedures are prisoners of the data: if the data are biased, so too will the results be. This illustrates the imperative that research unequivocally isolates variables, measures them with valid and reliable tools, and restricts the interpretation of outcomes to those that are core to the research question.

Key words: Meta-Analysis, Computer Technology, Higher Education, Effect Size, Achievement.