Communications of the Association for Information Systems

Volume 16 Article 32

October 2005

Understanding the Role and Methods of Meta-Analysis in IS Research

William R. King
University of Pittsburgh, billking@katz.pitt.edu

Jun He
University of Pittsburgh, JunHe@katz.pitt.edu

Communications of the Association for Information Systems

UNDERSTANDING THE ROLE AND METHODS OF META-ANALYSIS IN IS RESEARCH

William R. King
Jun He
Katz Graduate School of Business
University of Pittsburgh
bill.king@katz.pitt.edu

ABSTRACT

Four methods for reviewing a body of research literature – narrative review, descriptive review, vote-counting, and meta-analysis – are compared. Meta-analysis as a formalized, systematic review method is discussed in detail in terms of its history, current status, advantages, common analytic methods, and recent developments.

Meta-analysis is found to be underutilized in IS. Suggestions on encouraging the use of meta-analysis in IS research and procedures recommended for meta-analysis are also provided.

Keywords: literature reviews, narrative review, descriptive review, vote counting, meta-analysis

I. INTRODUCTION

Compared with many other disciplines in business and the social sciences, Information Systems (IS) is a relatively young field. In its initial development stage, IS was perceived as an applied discipline that almost exclusively drew on other, more fundamental, reference disciplines [Keen, 1980]; its progress and maturation became a preoccupation to some researchers [Banville and Landry, 1989]. Now, IS "fully emerged as a discipline in its own right" [Baskerville et al., 2002; p. 1].

The development of IS as a scientific field is evidenced by the building of solid research tradition. In an empirical study of the disciplines cited in IS journals, Vessey et al. [2002] found a substantial volume of IS research used IS itself as the reference discipline (27% in their analysis). A survey [Palvia et al., 2004] of the use of different research methodologies in IS showed that about 6.6% of studies published in seven IS journals during the years 1993-2003 used literature reviews as the primary research methodology.

IS is also emerging as an important reference discipline for other fields [Baskerville et al., 2002]. A search of the Social Science Citation Index (SSCI) showed that it is common for many fields to conduct research based on IS theories or models. A search with the keyword of "information systems or information technology" resulted in a total of 719 articles published in 329 refereed journals in 2004. Of these journals, 26 are IS-focused, while the other 303 journals are from various fields covering almost all the research areas of business and social sciences. A good example of this use of IS theory in other areas is the technology acceptance model (TAM), which was tested and applied in psychology [e.g., Gentry et al., 2002], education [e.g., Liaw 2002;

Selim, 2003], marketing [e.g.,Keen et al., 2004], operations management [e.g., Olson et al., 2003], and many other management fields [King and He, 2005].

However, IS research inevitably displays one problem that is common in many mature fields: inconsistent empirical findings on essentially the same question. Knowledge accumulation increasingly relies on the integration of previous studies and findings. As Glass [1976] suggested, when the literature on a topic grows and knowledge lies untapped in completed research studies,

> *"this endeavor (of research synthesis) deserves higher priority .... than adding a new experiment or survey to the pile"* [Glass, 1976; p. 4].

A recent trend in research synthesis is to integrate, quantitatively, knowledge garnered from empirical studies on a topic by using meta-analysis. Meta-analysis is the most commonly used quantitative research synthesis method in the social and behavioral sciences [Hedges and Olkin, 1985]. It won recognition as a better way to conduct reviews of a body of completed research than the traditional narrative fashion [Wolf, 1986; Hunter and Schmidt, 1990; Rosenthal, 1991; Cooper and Hedges, 1994; Rosenthal and DiMatteo, 2001]. Some journals revised their review policy to encourage the use of this methodology [e.g., From the Editors of *Academy of Management Journal*, 2002]. In IS, however, meta-analysis is extremely underutilized [Hwang, 1996]. Moreover, some of the meta-analytic practices used in meta-analysis in IS are conceptually or methodologically flawed [King and He, 2005].

The objective of this paper is to provide guidelines for IS researchers who are interested in synthesizing a body of literature in a rigorous and quantitative fashion. After reviewing and comparing different literature synthesis methods (Section II), we turn our attention to meta-analysis. We describe the history, common methods, and recent developments of meta-analysis (Sections III and IV). We also provide a discussion of some major concerns with existing meta-analysis applications (Section V) and the great potential of applying meta-analysis in IS (Section VI).

II. LITERATURE REVIEW METHODS AND META-ANALYSIS

The refinement and accumulation of information and knowledge are an essential condition for a field to "be scientific" and to progress [Hunter et al., 1982; Pillemer and Light, 1980]. Researchers can use a number of techniques for making sense out of existing research literature, all with the purpose of casting current research findings into historical contexts or explaining contradictions that might exist among a set of primary research studies conducted on the same topic [Rumrill and Fitzgerald, 2001]. Many researchers dichotomize literature review methods as qualitative versus quantitative reviews [e.g., Wolf, 1986; Aldag and Stearns, 1988; Hunter and Schmidt, 1990; Rosenthal and DiMatteo, 2001; and Palvia et al., 2003]. This approach may be overly simplistic in that different review techniques vary in the extent of systematically synthesizing an existing literature body, ranging from purely qualitative (e.g., verbal description) to moderately quantitative (e.g., counting a number or calculating a percentage of certain research characteristic) to purely quantitative (e.g., meta-analysis).

Following Guzzo et al.'s [1987] approach, we categorize the most commonly employed review techniques in IS along a continuum of quantification as narrative reviews, descriptive reviews, vote counting, and meta-analysis (Figure 1).

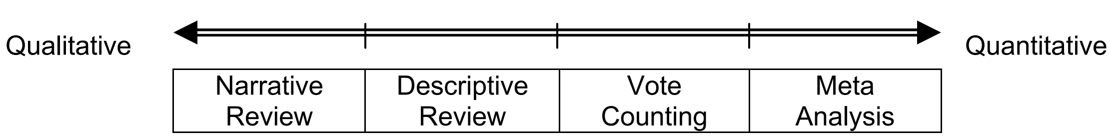


Figure1. Review Methods on a Qualitative-Quantitative Continuum

NARRATIVE REVIEWS

Narrative reviews present verbal descriptions of past studies focusing on theories and frameworks, elementary factors and their roles (predictor, moderator, or mediator), and/or research outcomes, (e.g., supported vs. unsupported) regarding a hypothesized relationship. Narrative reviews are of great heuristic value, and serve to postulate or advance new theories and models, to examine important and/or controversial topics, and to direct further development in a research domain.

No commonly accepted or standardized procedure for conducting a narrative review exists. This lack is a key weakness of narrative reviews as a means of arriving at a firm understanding of a research tradition [Green and Hall, 1984; Hunter and Schmidt, 1990; Rosenthal and DiMatteo, 2001]. Researchers are relatively free to design their review strategy in terms of selecting relevant papers, categorizing research characteristics, and framing outcomes.

When conducting a narrative review, researchers tend to consciously or subconsciously make judgments that support their own background, understanding, or established point-of-view. Often their goal is to come to some conclusions through classifications of the research methods and categorizations of results. One commonly-used strategy is to create a classification or a typology to organize the results. Researchers adopting this approach should be cautious that the creation of categories may result in less information since quantitative scales may be reduced to qualitative ones. Of course, this approach can result in greater understanding and lead to greater insight, but such a result is certainly not assured.

Ives and Olson's [1984] study on user involvement may serve as a representative narrative review in IS. Ives and Olson proposed a model of user involvement prior to their review of the literature; and this model framed their discussions of the selected papers. Ives and Olson's [1984] work is considered influential in the field of user involvement.

DESCRIPTIVE REVIEWS

Descriptive reviews introduce some quantification, often a frequency analysis of a body of research. The purpose is to find out to what extent the existing literature supports a particular proposition or reveals an interpretable pattern [Guzzo et al., 1987].

To assure the generalizability of the results, a descriptive review often involves a systematic search of as many as relevant papers in an investigated area, and codes each selected paper on certain research characteristics, such as publication time, research methodology, main approach, grounded theory, and symbolic research outcomes (e.g., positive, negative, or non-significant). A frequency analysis (including its derivatives of trend analysis and cluster analysis) treats an individual study as one data record and identifies distinct patterns among the papers surveyed. In doing so, a descriptive review may claim its findings to represent the fact or state of a research domain.

Palvia et al.'s [2003; 2004] analyses of the use of research methodologies in IS are typical descriptive reviews. In the two studies, Palvia and colleagues surveyed articles in seven IS journals (communications of the ACM, Decision Sciences, Information & Management, Information Systems Research, Journal of Management Information Systems, and Management Science) in the years 1993 to 2003. They coded each article for up to two methodologies (primary methodology and secondary methodology), and calculated the frequency and analyzed the trend of each of thirteen research methodologies as used in these papers. The results "provide the current state of research methodologies in use" [Palvia et al., 2004; p. 306].

VOTE COUNTING

Vote counting, also called "combining probabilities" [Rosenthal, 1991] and "box score review" [Guzzo et al., 1987], is commonly used for drawing qualitative inferences about a focal relationship (e.g., a correlation is significantly different from 0 or not) by combining individual research outcomes [Pickard et al., 1998]. Some researchers consider vote counting a meta-analytic technology [e.g., Rosenthal, 1978; 1991]; and other researchers separate vote counting as an alternative quantitative review method mostly because this method does not analyze effect sizes. It uses the outcomes of tests of hypothesis reported in individual studies, such as probabilities, *p*-levels, or results falling into three categories: significantly positive effect, significantly negative effect, and non-significant effect. The philosophy is that repeated results in the same direction across multiple studies, even when some are non-significant, may be more powerful evidence than a single significant result [Rosenthal and DiMatteo, 2001].

Rosenthal [1991] provided a comprehensive review of nine different vote counting methods, and discussed the advantages and limitations of each. For illustrative purposes, in Appendix 1 we provide the computational formulas of two most popular vote counting methods: Fisher's procedure of combining p's and Stouffer's procedure of combining Z's.

Vote counting does not require other statistics such as effect sizes and construct reliabilities. Thus, it is a conceptually simple and practically convenient method. Under certain circumstances (i.e., small number of sampled studies, when investigated effects are in the same direction) this method could produce statistically powerful results. In one extreme example, Cohen [1993] reported on two studies that dealt with the results of vaccinating monkeys. Because laboratory animals are difficult to obtain and expensive to maintain, the studies involved only six and eleven monkeys respectively (both experimental subjects and control animals). Neither study produced statistical significance. However, when the data were combined in a vote counting analysis, the p-level was considerably smaller and the effect was shown to be large.

Vote counting contains some inherent limitations. In particular, it allows a weak test of a hypothesis (e.g., correlation is 0 or not 0) with little consideration of the magnitude of the effect, such as an estimated effect size and associated confidence intervals. In addition, vote-counting assumes homogeneity in the sample population investigated. In case of heterogeneity, which is most common in research, vote-counting cannot detect moderator effects, and the combined significance could be meaningless. Therefore, vote-counting is often suggested as a supplement to meta-analysis in the case of missing effect sizes [Bushman and Wang, 1995; Pickard et al., 1998].

In IS, vote counting is applied to produce a single quantitatively synthesized conclusion from a series experiments. In some fields within IS, such as software engineering, where much research involves modest experimental effects, small sample sizes, and hypothesis testing fails to conclude significant results due to low statistic power, vote counting is found to be particularly useful. For example, in a study of reading techniques for defect detection of software codes, Laitenberger et al. [2001] found that perspective-based reading (PBR) was statistically more effective than checklist-based reading (CBR) in one out of three experiments; the other two were in the same direction but not significant (p-value (one side) = 0.16 and 0.13). When applying Fisher's procedure, Laitenberger and colleagues concluded that PBR was a significantly more effective reading technique for defect detection than CBR, with a combined p-value = 0.000016 (p. 403). Similarly, Pfahl et al. [2004] applied vote counting to analyze results from a series of three experiments to assess the learning effectiveness of using a process simulation model for educating computer science students in software project management. The results supported that simulation involves more learning interests of students than that of without-simulation students (p-values from the 3 experiments were 0.04, 0.21, and 0.28; the combined p-value were 0.06 (Fisher's procedure) and 0.03 (Stouffer's Z procedure) (p. 137-138).

META-ANALYSIS

Meta-analysis is a statistical synthesis method that provides the opportunity to view the "whole picture" in a research context by combining and analyzing the quantitative results of many empirical studies [Glass, 1976]. It connotes a rigorous alternative to the narrative discussion of research studies which typify our attempts to make sense of the rapidly expanding research literature [Wolf, 1986]. Since we believe that meta-analysis is of great potential significance and find that it is often inadequately applied in IS, we devote the remainder of this paper to it.

III. OVERVIEW OF META-ANALYSIS

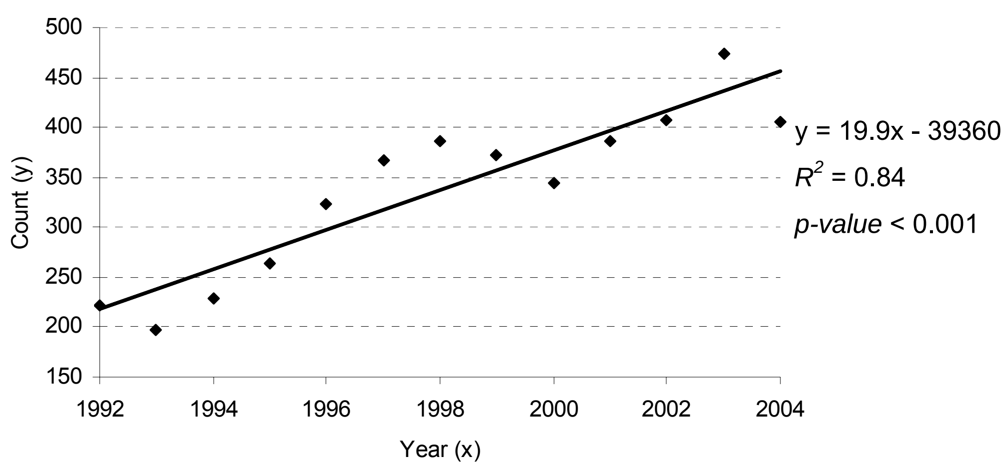
The history of the use of statistical methods to synthesize research findings is as long as most commonly-used statistical procedures [Cooper, 1979]. Hedges [1992] credited Legendre, who invented the principle of least squares to solve regression problems, as an innovator in this area because his original purpose was to combine information across studies. Another often-cited example is Pearson's [1904] study of enteric fever inoculation, in which Pearson presented sets of data collected under different conditions, calculated individual effect sizes that were comparable across data sets, examined homogeneity, and computed an average of the effect sizes to determine the effectiveness of the treatment (inoculation).

However, applications of this quantitative research synthesis method were rare before the 1970s [Cooper and Hedges, 1994]. It seems that the method grew in popularity since Glass coined the phrase "meta-analysis" in 1976.

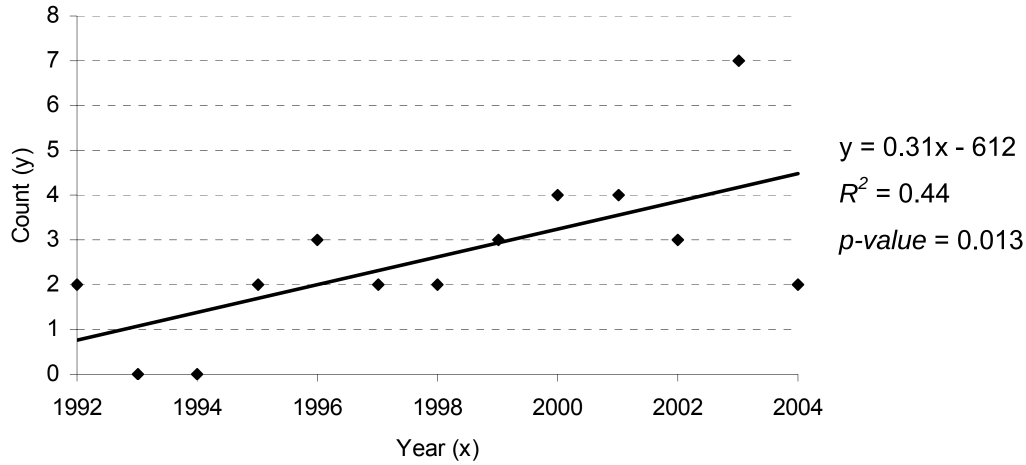
> *Meta-analysis refers to the analysis of analyses . . . the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding research literature* [Glass, 1976, p.3]

Meta-analysis is now a legitimate research review tool accepted by, and dominant in, many fields in the social and behavioral sciences [Field, 2001]. We searched a comprehensive social science database – Social Science Citation Index (SSCI) – with the keyword "meta-analysis". The search resulted in 4407 articles published in refereed journals in all covered fields for the period 1992-2004, with an average annual increase of 20 articles (p-value < 0.0001).

In IS, some researchers argued that meta-analysis is a rarely applied research methodology [e.g., Hwang, 1996; Palvia et al., 2003], mostly because of the small number of applications. For example, Hwang (1996) reviewed the use of meta-analysis in IS prior to 1996 and found 6 applications, including one conference paper. We searched SSCI with the restriction of IS as the research domain. The search resulted in 34 published meta-analyses in 1992-2004, with an average annual increase of 0.3 article (p-value = 0.013). The numbers of published meta-analyses in social science in general and in IS in particular are presented in Figure 2.



a. Social Sciences



b. Management Information Systems

Figure 2. Number of Meta-analyses Published in Refereed Journals

As evidenced in Figure 2b, the use of meta-analysis increased steadily in IS, although the numbers of applications are still small. We believe this trend will be maintained, given the maturating of the field and the growing number of research topics that would benefit from conclusive answers.

ADVANTAGES OF META-ANALYSIS

Meta-analysis is much less judgmental and subjective than other literature review methods, particularly narrative review, and therefore much more in tune with positivist tradition. The major difference between narrative reviews and quantitative meta-analyses may well be that narrative reviews primarily focus on the conclusions reached in various studies, whereas meta-analyses focus on data, as reflected by the operationalization of variables, the magnitude of effect sizes, and the sample sizes.

Qualitative assessments involved in a narrative reviews usually do not take account of both the relative sizes of samples and the significance of the effects measured. For example, a small-sample study with significant results may be equally weighted with a similar large-sample study in such assessments, while an insignificant result may be ignored when it does, in fact, contribute to a body of research that overall, may show significant effects.

To synthesize a research literature, statistical meta-analysis uses final results collected from a collection of similar studies [Glass, 1976]. The final results are effect sizes, or the magnitude of the effects. The focus on effect sizes rather than significances of these empirical findings is an advantage over traditional literature review methods.

Meta-analysis enables the combining of various results, taking into account the relative sample and effect sizes, thereby permitting studies showing insignificant effects to be analyzed along with others that may show significant effects. The overall result may be either significant or insignificant, but it is undoubtedly more accurate and more credible because of the overarching span of such an analysis.

Meta-analysis can focus attention on the cumulative impact of insignificant results that can be significant. For example, two studies showing sgnificance at the 0.06 level are much stronger evidence against the null hypothesis that is a single study at the 0.05 level, all else being equal. Yet, in some fields, the former two studies may not ever be published. Meta-analysis, by combining such results, enables us to see the big picture of the landscape of research results.

In other fields, meta-analysis has provided answersto questions that were in great dispute because of conflicts in the results of various studies. For example, the viewing of violence on television was shown to be associated with a greater tendency toward aggressive and anti-social acts through a meta-analysis of more than 200 studies, many of which were individually inconclusive or had reached contrary conclusions [Comstock et al., 1991].

In business research, perhaps the area which made the greatest use of meta-analysis is organizational behavior. There, meta-analyses seem to have been conducted primarily in the source of job performance ratings involving combinations of ratings by supervisor, subordinates, peers and self-ratings [Conway and Huffcutt, 1997; Harris et al., 1988].

While meta-analysis is basically confirmatory in nature, it can also involve exploratory aspects. For example, when high variability exists in the effects that are reflected in various studies, meta-analysis promotes a search for moderator variables. While this use of meta-analysis introduces greater subjectivity on the part of the researchers, the subjectivity is certainly less than that involved in performing a non-quantitative literature review.

CONCERNS WITH META –ANALYSIS

Numerous researchers advocated the use of meta-analysis as a better way to conduct research reviews [e.g., Glass, 1976; Hedges and Olkin, 1985; Wolf, 1986; Hunter and Schmidt, 1990; Rosenthal, 1991; Rosenthal and DiMatteo, 2001]. However, as with other research methods, meta-analysis is not free from limitations. In this subsection, we discuss some major concerns with meta-analysis as a research review method.

Sampling Bias toward Empirical Studies

By its statistic analysis nature, meta-analysis contains an inherent sampling bias toward quantitative studies that report effect sizes. Using Palvia et al.'s [2003; 2004] typology, the types of research that may provide data to a meta-analysis include survey, laboratory experiment, field study, and field experiment. Other types of research, such as frameworks and conceptual models, case studies, speculation/commentary, mathematical models, secondary data, interview, and qualitative research, are unlikely to be sampled in a meta-analysis. This limitation suggests that a significant number of studies (according to Palvia et al. [2004], these studies add up to 50.3% of all IS publications in seven leading IS journals) will be ignored when conducting a meta-analysis.

In addition, some researchers may combine quantitative and qualitative research methods in their studies of the same phenomenon, called triangulation [Gable, 1994; Kaplan and Duchon, 1988; Lee, 1991; Mingers, 2001]. Triangulation may help deepen and widen our understanding on a certain phenomenon. One example is Markus' [1994] study of the use of e-mail in organizations. Markus conducted a field survey and data did not provide much support to the hypotheses developed from information richness theory; Markus also interviewed some respondents, whose answers and comments provided insight on factors affecting their use of email. If a meta-analysis on communication effectiveness were to be conducted, only the empirical part of Markus's [1994] study would be included and the qualitative part, which provides more value to media use research, would be ignored.

The sampling bias toward empirical studies, a limitation of meta-analysis, is rarely addressed in discussions of this research method. We believe that it is important to call to the attention of meta analysts and researchers that the results from a meta-analysis are not necessarily more creditable or representative of a research population than those from a narrative review.

Garbage In and Garbage Out

Meta-analysis does not generally differentiate studies by their quality. This issue is often referred to as “garbage in and garbage out” [Hunt, 1997]. Research studies vary considerably in their research designs and approaches, sampling units, methods of measuring variables, data analysis methods, and presentations of research findings. The inclusion of the results from poorly-conducted studies with flawed designs into a meta-analysis could confuse the full understanding of the literature body investigated or even lead to unfounded conclusions. The judgment of quality is rather subjective, although some techniques have been suggested to correct this error [e.g., Rosenthal, 1991]. However, these techniques introduce other biases about the selection and weighting of quality criteria.

Publication Bias

Publication bias [Begg and Berlin, 1988], also known as the file-drawer problem [Rosenthal, 1979; Iyengar and Greenhouse, 1988], refers to the observation that significant results are more likely to be published while non-significant results tend to be relegated to file drawers. Thus, the meta-analysis result will focus on an unrepresentative proportion of a total research population. Of course, publication bias applies to all review methods. It is “particularly problematic for a meta-analysis whose data come solely from the published scientific literature” [Duval and Tweedie, 2000]. Using an unrepresentative set of data may result in conclusions biased toward significance or positivity.

Although some correction techniques have been developed [e.g., Duval and Tweedie, 2000], the best solution to avoid this bias is to search multiple databases in a systematic way and sample studies from various sources [Rosenthal and DiMatteo, 2001]. For example, in their study of the effects of management support and task interdependence on IS implementation, Sharma and Yetton [2003] searched the literature in various ways, including bibliographic databases, manual searches, and the bibliographies of existing works. They located 22 empirical studies, of which 11 are from journal publications, 9 from dissertations, and 2 from book chapters. The comprehensive search resulted in a diverse sample that “both increases the power of the meta-analysis by maximizing the number of studies and reduces (publication) source bias” (p. 542).

Apples And Oranges

One criticism of meta-analysis is that it may compare “apples and oranges,” aggregating results derived from studies with incommensurable research goals, measures, and procedures. It is argued, therefore, that meta-analysis may sometimes be analogous to taking apples and oranges and averaging such measures as their weights, sizes, flavors, and shelf lives [Hunt, 1997]. This problem exists for all review methods, qualitative or quantitative, in that “exact replications probably cannot occur” [Hall et al., 1994; p. 20], “(studies) ... are rarely precisely the same” [Rosenthal and DiMatteo, 2001; p. 68]. This problem is not of dominant significance, especially when we want results that are generalizable to fruits, or to a broad research domain. On the other hand, synthesists must be sensitive to the problem of attempting aggregation of too diverse a sampling of studies.

Small Sample Size

The statistical power of detecting a genuine effect size depends on both the number of studies and the total cumulated sample sizes included in a meta-analysis. The more studies that are included in the meta-analysis, the more creditable are the results at representing the investigated research domain. Using a Monte Carlo simulation, Field [2001] found that a meta-analysis should include at least 15 studies, otherwise the type I error (accepting a false null hypothesis) could be severely inflated.

If the investigator cannot identify enough empirical studies on a common topic, it may indicate that the research domain is too immature for a conclusive study such as meta-analysis.

> "*when meta-analytical results have less evidential value because the number of individual studies in the meta-analysis is small … the alternative is neither reversion to reliance on the single study nor a return to the narrative review method of integrating study findings: both are vastly inferior to meta-analysis in the information yield"* [Hunter and Schmidt, 1990; p. 420].

Reviewers should base analyses on the largest number of studies available [Matt and Cook, 1994]. If the collection of studies is largely incomplete, then conclusions drawn from analysis are limited in scope.

An exception will be meta analyzing a series of experimental studies. Here a researcher is interested in concluding generalizability of a proposition within, rather than beyond, a defined set of studies. For example, Laitenberger et al. [2001] and Phafl et al. [2004] meta analyzed a series of three experiments (an initial experiment and two replications). The combined results were stronger in statistical power than results calculated from any individual experiment.

Methodological Errors

Comparatively, meta-analysis is a straightforward or "formalized systematic review" procedure that is more standardized than other review methods [Hunter and Schmidt, 1990]. Given the same set of data or sampled effect sizes from a literature body, different researchers should arrive at the same conclusion via meta-analysis. In other words, a typical meta-analysis is often based on the procedures and analytic methods that are commonly accepted. Thus, the results from a meta-analysis are often treated as reliable and objective [Rosenthal and DiMatteo, 2001]. However, a particular meta-analysis may contain methodological errors that lead to a false conclusion.

We observed serious methodological errors in some IS meta-analyses. For example, in the two meta-analyses conducted by Mahmood and colleagues [2000, 2001], some combined effect sizes were larger than any sampled individual effect size. In their meta-analysis of the correlations between perceived ease of use and Information technology usage (Table 1 of Mahmood et al. [2001], p. 116), Mahmood and colleagues concluded a combined effect size as large as 0.678, while the sampled correlation coefficients ranged from 0.059 to 0.375. If the result is used by a researcher who performs statistical power analysis at the 80% level to guide a research design (suggested by Cohen [1988, 1992]), the researcher will conclude that a sample size of 28 may be adequate to detect a significant relationship (at α=0.05 level) between the two variables. In fact, the sampled studies in Mahmood et al. [2001] involved much larger sample sizes, ranging from 61 to 786, with a mean of 185.

Statistics Used in a Meta-Analysis. To meta-analyze an issue, the researcher should extract the following statistics and information from the studies:

- effect sizes
- sample sizes
- possible moderators
- construct reliabilities
- other measurement error indexes

Most IS studies report these statistics in their presentations.

In the literature, effect sizes are of various forms and can be categorized into two main families, the *r* family and the *d* family. The *r* family effect sizes report correlations between variables; the specific type depends on whether the variables are in continuous (Pearson's *r*), dichotomous (*phi*), or ranked form (*rho*). In IS research, the most popularly reported effect size is the Pearson's *r*. The *d* family are used mostly in laboratory experiments and measure the standardized difference between an experimental group and a control group. (There are three main members in this family: Cohen's *d*, Hedges' *g*, and Glass's Δ. The three are calculated in a similar way: the difference between two means is divided by some variance. They differ in the denominator: the square root of the pooled variance of the two groups for Cohen's *d*, the square root of the pooled variance for Hedges' *g*, and the square root of the control group variance for Glass's Δ.) These effect sizes can be readily converted to one another. Effect sizes can also be calculated from various testing statistics, such as *t, F*, χ^2, or *Z*, or their associated *p* levels. Detailed computational formulas to calculate and convert these statistics are beyond the purpose and scope of this study. Interested readers can refer to Wolf [1986], Hunter and Schmidt [1990], Rosenthal [1991], and Rosenthal and DiMatteo [2001].

Other statistics, including the descriptive statistics of investigated variables, cannot be directly used as effect sizes. As pointed out by Rosenthal and Dimatteo [2001], "an r effect size cannot be computed from kappa, percent agreement, relative risk, risk difference, or the odds ratio unless all the raw data are available so the meta-analyst can compute the proper index" (p.72).

Fixed vs. Random Effects Models, Similar to other statistical methods, meta-analysis methods are developed based on assumptions about the population from which studies are taken. The two common assumptions lead to two different analysis methods: fixed-effects and random-effects models.

The fixed-effects model assumes that studies in the meta-analysis are sampled from one population with a fixed "true" effect size. In other words, the true effect size is assumed to be the same for all studies included in the analysis, and the observed variance among effect sizes is dominated by sampling errors, which are unsystematic and can be estimated [Hunter and Schmidt, 1990]. The assumption of one population underlying a meta-analysis restricts the conclusions from being generalized to a study not included in the analysis unless the study shows independent evidence of belonging to the population; i.e., unless it is a close replication of the studies included in the meta-analysis.

In contrast, the random effects model assumes that population effect sizes vary from study to study. As such, a study included in such a meta-analysis can be viewed as being sampled from a universe of possible effects in a research domain [Field, 2001]. As long as the meta-analysis covers the literature comprehensively, the conclusions are generalizable to the research domain as a whole and can be applied to studies not included in the analysis.

In statistical terms the two models differ in the calculation of the weights used in the analysis, which in turn affects the standard errors associated with the combined effect size. Fixed-effects models use only within-study variability in their weights because all other "unknowns" in the model are assumed to be constant. In contrast, random-effects models account for the errors associated with sampling from populations that themselves were sampled from a superpopulation. The error term, therefore, contains variability arising from differences between studies in addition to within-study variability. Standard errors in the random-effects model are, therefore, larger than in the fixed case, which makes significance tests of combined effects more conservative [Field, 2003; p. 107].

Although random-effects models generally appear to be superior, fixed-effects models are in common use. If the fixed-effects model is employed for a meta-analysis, the assumption of one "true" effect size across studies should be tested before combining effect sizes [Hedges and Olkin, 1985]. The test of this assumption, labeled as a homogeneity test, is a chi-square test of the null hypothesis that all effect sizes are the same after controlling for sampling errors. The test result indicates whether the null hypothesis can be rejected at a certain level. Only when the test result is insignificant can the sampled studies be combined (e.g., calculating the combined effect sizes and their confidence intervals). In many cases, the test indicates a violation of the assumption and there is a need to switch to the random-effects model; however, in practice sometimes the fixed-effects model is chosen and the test is not performed.

In contrast, the random-effects model assumes variation between the populations underlying the various studies that are incorporated into the meta-analysis. The homogeneity test in this case examines whether the interaction between sampling error and between-study variance is zero or not. An insignificant test result (which occurs much of the time) justifies the techniques that are used under this model. As suggested by many meta-analysts [e.g., Field, 2001; Field, 2003; Hedges and Vevea, 1998; Rosenthal and DiMatteo, 2001], homogeneity is rare and the random-effects model should be applied in most research domains.

PRIMARY META-ANALYSIS METHODS

Over the past 20 years, three methods of meta-analysis emerged to be widely used [Field, 2001; Johnson et al., 1995]: the method devised

- by Hedges and Olkin [1985];
- by Rosenthal and Rubin [Rosenthal, 1991], and
- by Hunter and Schmidt [1990].

In IS, a fourth method ─ that devised by Glass and his associates [Glass et al., 1981] is also used. The four methods are briefly discussed subsection. Interested readers may examine the references for more detailed discussion on the methodologies. In addition,

- Johnson et al. [1995] and Field [2001] compared the first three methods basing on Monte Carlo simulation tests;
- Cooper and Hedges [1994] provided integrative reviews of the commonly-adopted meta-analytic approaches as well as computational formulas;
- Lipsey and Wilson [2001] described well the bolts and nuts of the whole process.

These studies will serve as good resources for potential meta-analysts.

Hedges-Olkin Method

This approach to meta-analysis is based on a weighted least squares technique [Hedges, 1987]. In this approach, study outcomes (i.e., effect sizes *r*) are transformed into a standard normal metric (using Fisher's *r*-to-*Z* transformation). Then, the transformed effect sizes are weighted by the inverse variances of each study. For the fixed-effects model, the variance is within-study variance, which is determined by sample sizes only. For the random-effects model, the variance is composed of within-study variance and between-study variance, the latter of which is from a chi-square test (Q test under the fixed-effects model). The combined effect size is the average of the weighted effect sizes, and its variance is the reciprocal of the sum of the weights. A significance test (*Z*-test) and confidence intervals of the combined effect size are then calculated.

Rosenthal-Rubin Method

Under a fixed-effects model, the Rosenthal-Rubin method employs essentially the same techniques as the Hedges-Olkin method, except for the significance test of combined effect size [Field, 2001]. Rosenthal and Rubin [Rosenthal, 1991] advocate the use of significance metrics (i.e., *Z*s associated with one-tailed probabilities) from sampled studies and examining the combined *Z* for the overall significance of the mean effect size.

Rosenthal [1991] did not present a random-effects version of his meta-analysis procedures in his original work. In a later study, Rosenthal and DiMatteo [2001] suggested "un-weighting" the effect sizes when meta-analytically integrating them as an approach based on a random-effects model. The basic logic is to treat studies as the unit of analysis in observing of the between-study variance. Consistent with the random effects model, the combined effect size of the unweighted-effect approach is less statistically powerful and has larger confidence intervals as contrasted with that calculated from the weighted-effect approach, and can be generalized to studies not yet in the sample. "If only one approach were to be used, it would be the one we prefer" [Rosenthal and DiMatteo, 2001; p. 70].

Hunter-Schmidt Method

This method is distinct for its efforts to correct effect size indices for potential sources of errors before meta-analytically integrating the effect sizes across studies [Johnson et al., 1995]. Besides the sampling error, other sources of errors include measurement error, range variation, construct validity, and variance due to extraneous factors. Errors other than sampling are labeled as "Attenuation Factors" in Hunter and Schmidt [1990], because they impact lower the observed correlations between investigated variables systematically, which can be corrected if assessed. Therefore, when information about these sources of error is available, this feature of Hunter-Schmidt method may recommend its use [Johnson et al., 1995]. In IS, at least one source of error, measurement error, is routinely assessed and reported by construct reliability (e.g., Cronbach α) [Chau 1999]. However, of the IS meta-analyses that we identified as using this method, none study explored this feature and corrected measurement error or other sources of error.

To test moderator variables, Hunter and Schmidt [1990] suggested a partitioning approach; that is, to break the set of studies into subsets using the moderator variable and to do separate meta-analyses within each subset of studies. Then, the difference between subsets is tested to conclude the magnitude and significance of this moderator variable.

Glass Method

Glass's approach to meta-analysis focuses on the detection of moderator variables. This method can be summarized as:

1. simulation of descriptive statistics across studies;
2. the coding of perhaps 50 to 100 study characteristics, such as date of study and number of threats to internal validity, and
3. regression of study outcome onto the coded study characteristics.

The characteristics that show significant effects on the study outcome are considered to be moderators.

Using a Monte Carlo test, Hunter and Schmidt [1990, p. 86-89] illustrated that this method has a severe problem of capitalization on (chance) sampling errors. Sampling errors are large because the sample size for looking at study characteristics is not the total number of subjects in the studies, but the number of studies. Correlating various study characteristics with study outcome leads to massive capitalization on chance when the correlations that are large enough to be statistically significant are identified ex post facto. A general suggestion for this approach is to derive few moderators basing on logical reasoning and existing theory before conducting the meta-analysis.

IV. RECENT DEVELOPMENTS IN META-ANALYSIS

After the term was coined by Glass in 1976, meta-analysis received much research attention. In early 1980s, many discussions of meta-analysis were centered on the legitimacy of this research method, i.e., comparing with other review methods, and identifying the advantages, limitations, and statistically soundness of meta-analysis [e.g., Glass, 1976; Hunter et al., 1982; Chow, 1987; Hedges, 1987]. The discussions of meta-analysis progressed to more advanced methodological topics, such as its application to structural equation modeling [Hom et al., 1992], and levels of analysis [Ostroff and Harrison, 1999]. This section reviews recent developments of meta-analysis in the analysis of moderator effects and mediator effects, the two issues that are most relevant to theory testing/building in IS.

ANALYSIS OF MODERATOR EFFECTS

In the context of meta-analysis, a moderator variable is a systematic difference among studies under review that might explain differences in the magnitudes or signs of observed effect sizes. The study of moderator effects often involves two stages: the detection, or exploratory analysis, of possible moderators, and the assessment, or confirmatory analysis, of theoretically suggested moderators.

The various moderator identification and assessment methods are shown in

Table 1. Common Methods of Detecting/Assessing Moderator Effects

Technology	Detecting moderator effects (exploratory)	Assessing moderator effects (confirmatory)
Graphing	X	
Q statistics (χ^2 test)	X	
Variance analysis	X	
Outlier test	X	
OLS regression	X, not suggested	X
WLS regression	X, not suggested	X
Partition test		X, not suggested for continuous moderator factors

Graphing. As sample size increases, the effect size would theoretically approach the population effect size. Sample effects sizes have greater variability with smaller samples than with larger samples. Simple and intuitive, graphing is suggested for preliminarily detecting naturally occurring groupings and possible moderator effects in a meta-analytic data set [Rosenthal and DiMatteo, 2001].

Q Statistics or Chi-Square Test [Hedges and Olkin, 1985]. This test generates a decision rule specifying whether the variability in standardized effect sizes is statistically significant. The test is based on a chi-square assessment of the level of variance across study results relative to the sampling error variance across studies. This test is often referred to as homogeneity test, and can serve as a criterion for selecting a fixed effects model vs. a random effects model for a meta-analysis.

Variance Analysis [Hunter et al., 1982]: The variance of sampled effect sizes is corrected for statistical artifacts, such as sampling error, differential reliability, and restriction of range. When artifacts fail to account for 75% of the variance, a search for moderator variables is indicated [Hunter and Schmidt, 1990].

Outlier Test. Studies that contain effect sizes more than two or three standard deviations from the mean may be examined. Differences and similarities between the studies in the tails may suggest possible moderator variables. Traditional outlier detection techniques include box-plot analysis for univariate data [Mosteller and Hoaglin, 1991] and studentized residuals for bivariate data [Freund and Littell, 1991]. An advance in this area is Huffcutt and Auther's [1995] Sample Adjusted Meta-Analytic Deviancy (SAMD) statistic, which takes into the account the sample sizes in a meta-analytic data set. It is based on the principle of sampling error that effect sizes based on a small sample size are more likely to be deviant than those from a large sample size [Hunter and Schmidt, 1990]. Beal et al. [2002] further discussed possible refinements of this technique by adjusting the inherent skewed distribution of effect sizes.

Ordinary Least Squares (OLS) [Glass, 1977]. Regress the effect size on the potential moderator variable, using individual study effect size as the dependent variable and the moderator variable as the independent variable [Wolf, 1986; Glass et al., 1981]. If the coefficient is significant, then the effect of the moderator variable may be significant. This method is criticized for risk of capitalization on chance [Hunter and Schmidt, 1990], and is suggested for assessing moderator effects only on a confirmatory basis.

Weighted Least Squares or WLS [Hedges and Olkin, 1985]. OLS assumes constant variance across observations or moderator variables retrieved from individual studies. This assumption is violated because sample error (main component of variance after controlling for moderator effects) is a function of effect size and sample size. Hedges and Olkin [1985] suggested using WLS and weight the multiple regression according the inverse of the sampling error variance. This method, although less popular than OLS, gives a more accurate assessment of moderator effects in most conditions [Steel and Kammeyer-Muller, 2002].

Partition Test [Hunter and Schmidt, 1990]. This test divides sampled effect sizes into subgroups by moderator factors, and compares subgroup means and variance, to assess if the means are significantly different. This method is particularly suggested for categorical moderator factors (e.g., gender, research methods, analysis levels, technology contexts). When applied to continuous moderator factors, such as by dividing data set into subgroups along a continuous moderator variable, this method performs poorly and underestimates moderator effects in most conditions [Steel and Kammeyer-Muller, 2002].

MEDIATOR ASSESSMENT METHODS

Meta-analysis was initially developed to examine first-order effects, such as treatment effects (the “d” group effect sizes) or correlation effects (the “r” group effect sizes). Other effects, such as mediator effects or partial correlation coefficients, were not addressed. Such relations can often form the basis of a theory and help establish a mediator mechanism or explain a causal relationship.

Assessing mediator effects in structural relationships as an important meta-analytic topic received much attention. Becker [1995] developed a technique to address structural relationships, analyzing whether a common population correlation matrix underlies a set of sampled empirical results. The analysis unit is correlation matrix instead of correlations. As few studies report the correlation matrix, application of the technique is limited in practice. One illustrative, but unsuccessful, example in IS is Legris et al.’s [2003] meta-analysis of the technology acceptance model (TAM). After an extensive search of empirical TAM studies, Legris and colleagues found usable matrices in only three out of 22 studies. Therefore, the small sample size resulted in “a statistic shortfall” and “limit(ed) the presentation of the findings to the general conclusion” (p. 202).

Two alternative approaches to study mediator effects are:

1. combining and analyzing meta-analytically-developed correlations; and
2. directly studying coefficients of interest as the effect sizes [Rosenthal and DiMatteo, 2001].

Taking TAM as an example, the core model (figure 3) suggests that perceived ease of use (EU) and perceived usefulness (U) are two important predictors of an individual’s behavioral intention to adopt a technology (BI); in addition, perceived usefulness partially mediates the effect of perceived ease of use on behavioral intention. The correlation coefficients (*r*’s) and

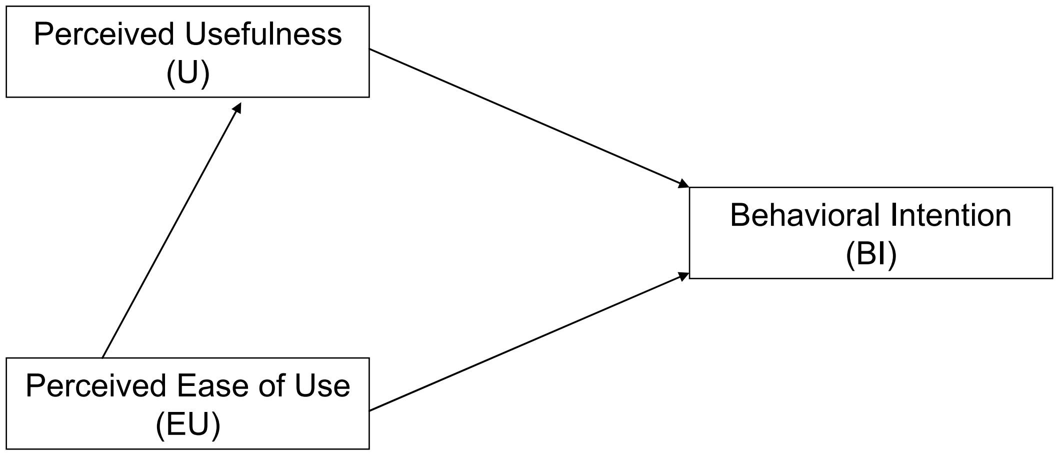


Figure 3. Technology Acceptance Model

pathcoefficients (βs) present the following relationship:

$$\beta(EU \rightarrow BI) = \frac{r_{(EU,BI)} - r_{(U,BI)} \times r_{(EU,U)}}{(1 - r^2_{(EU,U)})} \quad (1)$$

$$\beta(U \rightarrow BI) = \frac{r_{(U,BI)} - r_{(EU,BI)} \times r_{(EU,U)}}{(1 - r^2_{(EU,U)})} \quad (2)$$

$$\beta(EU \rightarrow U) = r_{(EU,U)} \quad (3)$$

The three equations hold for linear-regression-based analyses; they may differ a bit for structural-equation-modeling-based analyses (e.g., PLS, LISREL) because of different algorithms. But the differences are minor. In other words, we can infer the magnitude and the strength of path coefficients based on a set of meta-analytically-developed correlation coefficients. When applying the second approach – combining βs as the effect sizes, special caution should be taken that the sampled coefficients represent the relationship between the independent and the dependent variable controlling for other factors.

Both approaches were applied in another TAM meta-analysis conducted by King and He [2005]. This study meta-analyzed 88 TAM empirical studies, and the results from the two approaches were almost identical.

V. SUMMARY AND SUGGESTIONS

In this article, we first compare different review methods, then focus our discussion on meta-analysis, an important quantitative literature review method that is underutilized but which appears to offer great potential in IS literature synthesis research.

The benefits from meta-analysis changed the course of research in several fields, and it is possible that these benefits can be achieved in IS as well. One distinct feature of IS is the rapid changes that occurred in both the technologies and the applications. The dynamic nature of IS requires accurate assessments of newly developed technologies and business practices in a timely fashion. By synthesizing existing empirical studies, meta-analysis can serve as an efficient tool to satisfy the needs for overall conclusions about phenomena, without the burden of conducting new research in a particular situation, or the dilemma of selecting from competing theories or different perspectives.

Some researchers already work in this direction. For example, the development of various group support systems (GSS) in academia encouraged the use of teams, especially virtual teams, in organizational settings. However, the effectiveness of GSS was questioned by some inconsistent empirical findings in the literature [Fjermestad and Hiltz, 1999]. To address the concern, Dennis and Wixom [2003] meta-analyzed previous empirical studies. Their results not only validated GSS on improving group performance, but also explained conditions (i.e., the moderator effects of task, tool, the type of group, the size of the group, and facilitation) under which GSS would be most effective. Similar studies were done on computer-mediated-communication [Baltes et al., 2002], computer graphics [Hwang and Wu, 1990], and distance education [Allen et al., 2004]. Great potential exists for meta-analyzing other emerging or changing technologies and business practices, such as virtual teams and virtual organizations, knowledge acquisition technology (e.g., different interview techniques), system development practices (e.g., prototyping, user-centered system design, and rapid application development), IT governance, and knowledge management systems and practices, to name a few. A few empirical studies addressed these issues, and overall conclusive results will warrant the advance of the areas.

We expect meta-analysis will help direct future research in other IS issues. For example, the study of IT productivity (or payoff of IT investment) generated discussions of pros and cons of IT investment by both academicians and practitioners, a debate commonly known as the "productivity paradox" [Roach, 1987; Stassmann, 1985]. Kohli and Devaraj [2003] meta-analyzed existing literature on this issue. Their study not only validated the relationship between IT investment and firm performance, but also identified various factors that influence the likelihood for an investigation to find such a relationship. A future study in this area may benefit from Kohli and Devaraj [2003] by following their design strategy.

In addition, meta-analysis may help improve the publication practice in the IS field. Many studies that are performed with small sample sizes are never reported and many studies that produce non-significant results are rejected by journals. The development of a tradition of meta-analysis in IS would encourage the "publication" of such studies ─ perhaps on web sites and in electronic journals.

Various techniques are used to conduct meta-analysis; However, while no single technique is universally-agreed-upon as a way to perform such a study [Hall et al., 1995], the basic procedures for conducting a meta-analysis are well-understood [Rosenthal and DiMatteo, 2001].

Based on a review of the commonly-used meta-analysis methods, we suggest the following procedures for conducting a meta-analysis in IS. The procedure is illustrative and is designed to ensure that readers share a common understanding of statistical meta-analysis.

1. Define the research domain and the relationships of interest.
2. Collect studies in a systematic way. Try multiple databases with the same selection criterion not only to enlarge the data pool, but also to avoid bias toward certain journals.
3. Extract effect sizes ─ the strength of a relationship or the magnitude of a difference[1]. If the researchers did not reported the desired effect sizes, scour the articles for the

[1] In IS, the most popular form of effects is the Pearson product-moment correlation *r* ; standardized path coefficients can also be used, if all other factors are well controlled in a specific nomological network.

information necessary to calculate these effects. It is also recommended contacting the authors for the needed information[2].

4. Select either the random effects (suggested for most cases) or the fixed-effect model, and a specific analysis method to combine the effect sizes.
5. Examine the variability among the obtained effect sizes. Perform a homogeneity test, or plot effect sizes graphically to judge the range of effect sizes and the existence of possible moderating effects.
6. Examine the signs and magnitudes of the combined effect sizes (the means). Although t-tests are commonly employed for the significance level of an effect size, it is more useful to calculate confidence intervals around the mean to indicate the range of the effect size [Rosenthal and DiMatteo, 2001]. This approach is especially appropriate for the random-effects model, because this model assumes variations among effect sizes. A combined effect size implies no more than an average of a set of possible population effect sizes.
7. If the objective of the research is to find or test moderating effects, studies should be coded for characteristics of their contexts. Code the characteristics of individual articles in an unbiased way. Coders (often Ph.D. students or colleagues) should be “blind” to the research, and the internal consistency of the coding results should be tested.
8. To test the moderating effects, two methods are commonly used:
   a. Subgroup comparison: group studies according to their coded research contexts, then compare the combined effect sizes that are calculated within each subgroup.
   b. Regression: regress the research study characteristics on the effect sizes, the significant factors indicate the significant moderators.
9. Summarize and discuss the findings.

Editor’s Note: This article was fully peer reviewed. It was received on July 7, 2005 and was published on October 17, 2005.

REFERENCES

Ahituv, N., and R. Giladi (1993) “Business Success and Information Technology: Are they Really Related?” *Proceedings of 7th Annual Conference Management IS*, Tel Aviv University, Israel.

Aldag, R.J., and T.M. Stearns (1988) “Issues in Research Methodology”, *Journal of Management*, (14)2, pp. 253-276.

Banville, C.; and M. Landry (1989) “Can the Field of MIS be Disciplined?” *Communications of the ACM*, (32)1, pp. 48-60.

Baskerville, R.L., and M.D. Myers (2002) "Information Systems as a Reference Discipline", *MIS Quarterly*, (26)1, pp. 1-14.

Begg, C.B., and J.A. Berlin (1988) “Publication Bias: A Problem in Interpreting Medical Data”, *Journal of the Royal Statistical Society*, A, 151 (Part 3), pp. 419-463.

Bushman, B.J., and M.C. Wang (1995) “A Procedure for Combining Sample Correlation Coefficients and Vote Counts to Obtain an Estimate and a Confidence Interval for the

[2] We thank an anonymous reviewer of this study for this important recommendation.

Population Correlation Coefficient", *Quantitative Methods in Psychology*, (117)3, pp. 530-546.

Chau, P.Y.K. (1999) "On the Use of Construct Reliability in MIS Research: A Meta-Analysis", *Information & Management*, (35)4, pp. 217-227.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences, 2nd edition*, Hillsdale, NJ: Erlbaum

Cohen, J. (1992) "A Power Primer", *Psychological Bulletin*, (112)1, pp. 155-159.

Cohen, J. (1993) "A New Goal: Preventing Disease, Not Infection", *Science*, (262)5141, pp. 1820-1821.

Comstock, G., and H. Paik (1991) *Television and the American Child*, San Diego: Academic Press

Conway, J.M., and A.I. Huffcutt (1997) "Psychometric Properties of Multisource Performance Ratings: A meta-Analysis of Subordinate, Supervisor, Peer, and Self-Ratings", *Human Performance*, (10)4, pp. 331-361.

Cooper, H.M. (1979) "Statistical Combining Independent Studies: A Meta-analysis of Sex Differences in Conformity Research", *Journal of Personality and Social Psychology*, (37), pp. 131-146.

Cooper, H.M., and L.V. Hedges (eds.) (1994) *The Handbook of Research Synthesis*, New York: Russell Sage Foundation

Chow, S.L. (1987) "Meta-Analysis of Pragmatic and Theoretical Research: A Critique", *Journal of Psychology*, (121)3, pp. 259-271.

Duval, S., and R. Tweedie (2000) "A Nonparametric 'Trim and Fill' Method of Accounting for Publication Bias in Meta-Analysis", *Journal of the American Statistical Association*, (95)449, pp. 89-98.

Field, A.P. (2001) "Meta-Analysis of Correlation Coefficients: A Monte Carlo Comparison of Fixed- and Random-Effects Methods", *Psychological Methods*, (6)2, pp. 161-180.

Field, A.P. (2003) "The Problem in Using Fixed-Effects Models of Meta-Analysis on Real-World Data", *Understanding Statistics*, (2)2, pp. 105-124.

Fjermestad, J., and S.R. Hiltz (1999) "An Assessment of Group Support Systems Experimental Research: Methodology and Results", *Journal of Management Information Systems*, (15)3, pp. 7–149.

From the Editors (2002) "Replication, Meta-Analysis, Scientific Progress, and AMJ's Publication Policy", *Academy of Management Journal*, (45)5, pp. 841-846.

Gable, G. (1994) "Integrating Case Study and Survey Research Methods: An Example in Information Systems", *European Journal of Information Systems*, (3)2, pp. 112-126.

Gentry, L., and R. Calantone, (2002) "A Comparison of Three Models to Explain Shop-Bot Use on the Web", *Psychology & Marketing*, (19)11, pp. 945-956.

Glass, G.V. (1976) "Primary, Secondary, and Meta-Analysis of Research", *Review of Research in Education*, (5)10, pp. 351-379.

Glass, G.V., B. McGaw, and M.L. Smith (1981) *Meta-Analysis in Social Research*, Beverly Hills, CA: Sage

Green, B.F., and J.A. Hall (1984) "Quantitative Methods for Literature Reviews", *Annual Review of Psychology*, (35), pp. 37-53.

Guzzo, R.A., S.E., Jackson, and R.A. Katzell (1987) “Meta-Analysis Analysis”, *Research in Organizational Behavior*, (9), pp. 407-442.

Hall, J., and R. Rosenthal (1995) "Interpreting and Evaluating Meta-Analysis", *Evaluation & the Health Professions*, (18), pp. 393-407.

Hall, J.A., L. Tickle-Degnen, R. Rosenthal, and F. Mosteller (1994) “Hypotheses and Problems in Research Synthesis”, in Cooper, H. and L.V. Hedges (eds.) *The Handbook of Research Synthesis*, New York: Russell Sage Found, pp. 17-28.

Harris, M.E., and J. Schaubroeck (1988) "A Meta-Analysis of Self-Supervisor, Self-peer, and Peer-Supervisor Ratings", *Personnel Psychology*, (41)1, pp. 43-62.

Hedges, L.V. (1987) "How Hart IS Hard Science, How Soft is Soft Science", *American Psychologist*, (42)2, pp. 443-455.

Hedges, L.V. (1992) "Meta-Analysis", *Journal of Educational Statistics*, (17)4, pp. 279-296.

Hedges, L.V., and I. Olkin (1985) *Statistical Methods for Meta-Analysis*, New York: Academic

Hedges, L.V., and J.L. Vevea (1998) "Fixed- and Random-Effects Models in Meta-Analysis", *Psychological Methods*, (3)4, pp. 486-504.

Hom, P.W., F. Caranikas-Walker, G.E. Prussia, and R.W. Griffeth (1992) “A Meta-Analytical Structural Equations Analysis of a Model of Employee Turnover”, *Journal of Applied Psychology*, (77)6, 890–909.

Hunt, M. (1997) *How Science Takes Stock: The Story of Meta-Analysis*, New York: Russell Sage Foundation

Hunter, J.E., and F.L. Schmidt (1990) *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, Newbury Park, CA: Sage

Hunter, J.E., F.L. Schmidt, and G.B. Jackson (1982) *Meta-Analysis: Cumulating Research Findings Across Studies*, Beverly Hills: Sage

Hwang, M.I. (1996) “The Use of Meta-Analysis in MIS Research: Promises and Problems”, *the Data Base for Advances in Information Systems*, (27)3, pp. 35-48.

Ives, B., and M.H. Olson (1984) “User Involvement and MIS Success: A Review of Research”, *Management Science*, (30)5, pp. 586-603.

Iyengar, S., and D.B. Greenhouse (1988) “Selection Models and the File-Drawer Problem”, *Statistical Science*, (3), pp. 109-117.

Johnson, B.T., B. Mullen, and E. Salas (1995) “Comparison of Three Major Meta-Analytic Approaches”, *Journal of Applied Psychology*, (80)1, pp. 94-106.

Kaplan, B. and D. Duchon (1988) "Combining Qualitative and Quantitative Methods in Information Systems Research: A Case Study", *MIS Quarterly*, (12)4, pp. 571-587.

Keen, C., M. Wetzels, K. de Ruyter, and R. Feinberg (2004) "E-tailers versus Retailers: Which Factors Determine Consumer Preferences", *Journal of Business Research*, (57)7, pp. 685-695.

Keen, P.G.W. (1980) "MIS Research: Reference Disciplines and a Cumulative Tradition", in McLean, E. (eds.) *Proceedings of the First International Conference on Information Systems,* Philadelphia, PA, pp. 9-18.

King, W.R. and J. He (2005) “A Meta-Analysis of the Technology Acceptance Model”, Working Paper, University of Pittsburgh.

Kohli, R., and S. Devaraj (2003) “Measuring Information Technology Payoff: A Meta-Analysis of Structural Variables in Firm-Level Empirical Research“, *Information Systems Research*, (14)2, pp. 127-145.

Laitenberger, O., K. El Emam, and T.G. Harbich (2001) “An Internally Replicated Quasi-Experimental Comparison of Checklist and Perspective-Based Reading of Code Documents”, *IEEE Transactions on Software Engineering*, (27)5, pp. 387-421.

Lee, A. S. (1991) "Integrating Positivist and Interpretive Approaches to Organizational Research", *Organization Science,* (2)4, pp. 342-365.

Legris, P., J. Ingham, and P. Collerette (2003) "Why Do People Use Information Technology? A Critical Review of the Technology Acceptance Model", *Information & Management*, (40)3, pp. 191-204.

Liaw, S.S. (2002) "Understanding User Perceptions of World-Wide Web Environments", *Journal of Computer Assisted Learning*, (18)2, pp. 137-148.

Lipsey, M.W., and Wilson, D.B. (2001) *Practical Meta-analysis*, Thousand Oaks, CA: Sage Publications.

Mahmood, M.A., J.M. Burn, L.A. Gemoets, and C. Jacquez (2000) “Variables Affecting Information Technology End-User Satisfaction: A Meta-Analysis of the Empirical Literature”, *International Journal of Human-Computer Studies*, (52)4, pp. 751-771.

Mahmood, M.A., L. Hall, and D.L. Swanberg (2001) “Factors Affecting Information Technology Usage: A Meta-Analysis of the Empirical Literature”, *Journal of Organizational Computing and Electronic Commerce*, (11)2, pp. 107-130.

Markus, M.L. (1994) “Electronic Mail as the Medium of Managerial Choice”, *Organization Science*, (5)4, pp. 502-527.

Matt, G.E., and T.D. Cook (1994) “Threats to the Validity of Research Synthesis”, in Cooper, H. and L.V. Hedges (eds.) *The Handbook of Research Synthesis*, New York: Russell Sage Found, pp. 503-520.

Mingers, J. "Combining IS Research Methods: Towards a Pluralist Methodology", *Information Systems Research*, (12)3, pp. 240-259.

Olson, J.R., and K.K. Boyer (2003) "Factors Influencing the Utilization of Internet Purchasing in Small Organizations", *Journal of Operations Management*, (21)2, pp. 225-245.

Ostroff, C., and D.A. Harrison (1999) “Meta-Analysis, Level of Analysis, and Best Estimates of Population Correlations: Cautions for Interpreting Metaanalytic Results in Organizational Behavior”, *Journal of Applied Psychology*, (84)2, pp. 260–270.

Palvia, P., E. Mao, A.F. Salam, and K.S. Soliman (2003) "Management Information Systems Research: What's There in a Methodology", *Communications of the Association for Information Systems*, (11), pp. 289-308.

Palvia, P., E. Mao, V. Midha, P. Pinjani, and A.F. Salam (2004) “Research Methodologies in MIS: An Update”, *Communications of the Association for Information Systems*, (14), pp. 526-542.

Pearson, K. (1904) "Report on Certain Enteric Fever Inoculation Statistics”, *British Medical Journal*, (2), pp. 1243-1246.

Pfahl, D., O. laitenberger, G. Ruhe, J. Dorsch, and T. Krivobokova (2004) “Evaluating the Learning Effectiveness of Using Simulations in Software Project Management Education: Results from a Twice Replicated Experiment”, *Information and Software Technology*, (46)2, pp. 127-147.

Pickard, L. M., B.A. Kitchenham, and P.W. Jones (1998) “Combining Empirical Results in Software Engineering”, *Information & Software Technology*, (40)14, pp. 811-821.

Roach, S. (1987) America's Technology Dilemma: A Profile of the Information Economy. Special economic study, Morgan Stanley, New York.

Rosenthal, R. (1978) "Combining Results of Independent Studies", *Psychological Bulletin*, (85), pp. 185-193.

Rosenthal, R. (1979) "The File Drawer Problem and Tolerance for Null Results", *Psychological Bulletin*, (86), pp. 638-641.

Rosenthal, R. (1991) *Meta-analytic Procedures for Social Research*, *revised edition*, Beverly Hills, CA: Sage

Rosenthal, R., and M.R. DiMatteo (2001) "Meta-Analysis: Recent Developments in Quantitative Methods for Literature Reviews", *Annual Review of Psychology*, (52)1, pp. 59-82.

Rumrill, P.D. Jr., and S.M. Fitzgerald (2001) "Using Narrative Literature Reviews To Build A Scientific Knowledge Base", *Work*, (16)2, pp. 165-170.

Selim, H.M. (2003) "An Empirical Investigation of Student Acceptance of Course Websites", *Computers & Education*, (40)4, pp. 343-360.

Sharma, R., and P. Yetton (2003) "The Contingent Effects of Managemetn Support and Task Interdependence on Successful Information Systems Implementation", *MIS Quarterly*, (27)4, pp. 533-555.

Strassmann, Paul A. 1985. *Information Payoff: The Transformation of Work in the Electronic Age*. Collier Macmillan, New York, London: Free Press.

Vessey, I., V. Ramesh, and R.L. Glass (2002) "Research in Information Systems: An Empirical Study of Diversity in the Discipline and Its Journals", *Journal of Management Information Systems*, (19)2, pp. 129-174.

Wolf, F.M. (1986) *Meta-analysis: Quantitative Methods for Research Synthesis*. Beverly Hills, CA: Sage

APPENDIX I. VOTE COUNTING TECHNIQUES

Combining the p-values for the case with very small sample sizes can be performed either with the Fisher procedure, i.e. testing the statistic

$$P = -2\sum_{i=1}^{k} \ln(p_i)$$

which under H_0, corresponding to p_i; is χ^2-distributed with 2*k* degrees of freedom, or via calculating the p-value out of Stouffer's Z statistic:

$$Z = \frac{1}{\sqrt{k}}\sum_{i=1}^{k}(Z_i)$$

ABOUT THE AUTHORS

William R. King is University Professor in the Katz Graduate School of Business at the University of Pittsburgh. He is the author of more than 300 papers and 15 books in Information Systems, Management Science, and Strategic Planning. He was the Founding President of the Association for Information Systems, President of TIMS (now INFORMS) and Editor-in-Chief of the MIS Quarterly. He is a fellow of both AIS and INFORMS.

Jun He is a PhD Student in MIS at the Katz Graduate School of Business at the University of Pittsburgh. He presented papers at meetings of the Association for Computing Machinery (ACM) and the Americas' Conference on Information Systems (AMCIS) and also participated in the Doctoral Consortium at AMCIS.

Communications of the Association for Information Systems

ISSN: 1529-3181