

1

RESEARCH SYNTHESIS AS A SCIENTIFIC PROCESS

HARRIS COOPER
Duke University

LARRY V. HEDGES
Northwestern University

CONTENTS

1.1 Introduction	4
1.1.1 Replication and Research Synthesis	4
1.2 Research Synthesis in Context	4
1.2.1 A Definition of the Literature Review	4
1.2.2 A Definition of Research Synthesis	6
1.3 A Brief History of Research Synthesis as a Scientific Process	7
1.3.1 Early Developments	7
1.3.2 Research Synthesis Comes of Age	7
1.3.3 Rationale for the <i>Handbook</i>	10
1.4 Stages of Research Synthesis	11
1.4.1 Problem Formulation	11
1.4.2 Literature Search	12
1.4.3 Data Evaluation	12
1.4.4 Data Analysis	13
1.4.5 Interpretation of Results	14
1.4.6 Public Presentation	14
1.5 Chapters in Perspective	15
1.6 References	15

It is necessary, while formulating the problems of which in our advance we are to find the solutions, to call into council the views of those of our predecessors who have declared an opinion on the subject, in order that we may profit by whatever is sound in their suggestions and avoid their errors.

Aristotle, *De Anima*, Book 1, chapter 2

1.1 INTRODUCTION

The moment we are introduced to science we are told it is a cooperative, cumulative enterprise. Like the artisans who construct a building from blueprints, bricks, and mortar, scientists contribute to a common edifice, called knowledge. Theorists provide our blueprints and researchers collect the data that are our bricks.

To extend the analogy further yet, we might say that research synthesists are the bricklayers and hodcarriers of the science guild. It is their job to stack the bricks according to plan and apply the mortar that makes the whole thing stick.

Anyone who has attempted a research synthesis is entitled to a wry smile as the analogy continues. They know that several sets of theory-blueprints often exist, describing structures that vary in form and function, with few a priori criteria for selecting between them. They also know that our data-bricks are not all six-sided and right-angled. They come in a baffling array of sizes and shapes. Making them fit, securing them with mortar, and seeing whether the resulting construction looks anything like the blueprint is a challenge worthy of the most dedicated, inspired artisan.

1.1.1 Replication and Research Synthesis

Scientific literatures are cluttered with repeated studies of the same phenomena. Multiple studies on the same problem or hypothesis arise because investigators are unaware of what others are doing, because they are skeptical about the results of past investigations, or because they wish to extend (that is, generalize or search for influences on) previous findings. Experience has shown that even when considerable effort is made to achieve strict replication, results across studies are rarely identical at any high level of precision, even in the physical sciences (see Hedges 1987). No two bricks are exactly alike.

How should scientists proceed when results differ? First, it is clear how they should *not* proceed: they should not pretend that there is no problem, or decide that just one study (perhaps the most recent one, or the one they conducted, or a study chosen via some other equally arbitrary

criterion) produced the correct finding. If results that are expected to be very similar show variability, the scientific instinct should be to account for the variability by further systematic work.

1.2 RESEARCH SYNTHESIS IN CONTEXT

1.2.1 A Definition of the Literature Review

The American Psychological Association's *PsycINFO* reference database defines a literature review as "the process of conducting surveys of previously published material" (<http://gateway.uk.ovid.com>). Common to all definitions of literature reviews is the notion that they are "not based primarily on new facts and findings, but on publications containing such primary information, whereby the latter is digested, sifted, classified, simplified, and synthesized" (Mantel 1973, 75).

Table 1.1 presents a taxonomy of literature reviews that captures six distinctions used by literature review authors to describe their own work (Cooper 1988, 2003). The taxonomy was developed and can be applied to literature reviews appearing throughout a broad range of both the behavioral and physical sciences. The six features and their subordinate categories permit a rich level of distinction among works of synthesis.

The first distinction among literature reviews concerns the focus of the review, or, the material that is of central interest to the reviewer. Most literature reviews center on one or more of four areas: the findings of individual primary studies; the methods used to carry out research; theories meant to explain the same or related phenomena; and the practices, programs, or treatments being used in an applied context.

The second characteristic of a literature review is its goals. Goals concern what the preparers of the review hope to accomplish. The most frequent goal for a review is to *integrate* past literature that is believed to relate to a common topic. Integration includes: formulating general statements that characterize multiple specific instances (of research, methods, theories, or practices); resolving conflict between contradictory results, ideas, or statements

Table 1.1 A Taxonomy of Literature Reviews

Characteristic	Categories
Focus	research findings research methods theories practices or applications
Goal	integration generalization conflict resolution linguistic bridge-building criticism identification of central issues
Perspective	neutral representation espousal of position
Coverage	exhaustive exhaustive with selective citation representative central or pivotal
Organization	historical conceptual methodological
Audience	specialized scholars general scholars practitioners or policy makers general public

SOURCE: Cooper 1988. Reprinted with permission from Transaction Publishers.

of fact by proposing a new conception that accounts for the inconsistency; and bridging the gap between concepts or theories by creating a new, common linguistic framework.

Another goal for literature reviews can be to critically analyze the existing literature. Unlike a review that seeks to integrate existing work, a review that involves a critical assessment does not necessarily summate conclusions or compare the covered works one to another. Instead, it holds each work up against a criterion and finds it more or less acceptable. Most often, the criterion will include issues related to the methodological quality of empirical studies, the logical rigor, completeness or breadth of explanation if theories are involved, or comparison with the ideal treatment, when practices, policies or applications are involved.

A third goal that often motivates literature reviews is to identify issues central to a field. These issues may include questions that have given rise to past work, questions that

should stimulate future work, and methodological problems or problems in logic and conceptualization that have impeded progress within a topic area or field.

Of course, reviews more often than not have multiple goals. So, for example, it is rare to see the integration or critical examination of existing work without also seeing the identification of central issues for future endeavors.

A third characteristic that distinguishes among literature reviews, perspective, relates to whether the reviewers have an initial point of view that might influence the discussion of the literature. The endpoints on the continuum of perspective might be called neutral representation and espousal of a position. In the former, reviewers attempt to present all arguments or evidence for and against various interpretations of the problem. The presentation is meant to be as similar as possible to those that would be provided by the originators of the arguments or evidence. At the opposite extreme of perspective, the viewpoints of reviewers play an active role in how material is presented. The reviewers accumulate and synthesize the literature in the service of demonstrating the value of the particular point of view that they espouse. The reviewers muster arguments and evidence so that it presents their contentions in the most convincing manner.

Of course, reviewers attempting to achieve complete neutrality are likely doomed to failure. Further, reviewers who attempt to present all sides of an argument do not preclude themselves from ultimately taking a strong position based on the cumulative evidence. Similarly, reviewers can be thoughtful and fair while presenting conflicting evidence or opinions and still advocate for a particular interpretation.

The next characteristic, coverage, concerns the extent to which reviewers find and include relevant works in their paper. It is possible to distinguish at least four types of coverage. The first type, exhaustive coverage, suggests that the reviewers hope to be comprehensive in the presentation of the relevant work. An effort is made to include the entire literature and to base conclusions and discussions on this comprehensive information base. The second type of coverage also bases conclusions on entire literatures, but only a selection of works is actually described in the literature review. The authors choose a purposive sample of works to cite but claim that the inferences drawn are based on a more extensive literature. Third, some reviewers will present works that are broadly representative of many other works in a field. They hope to describe just a few exemplars that are descriptive of numerous other works. The reviewers discuss the characteristics that make

the chosen works paradigmatic of the larger group. In the final coverage strategy, reviewers concentrate on works that were highly original when they appeared and influenced the development of future efforts in the topic area. This may include materials that initiated a line of investigation or thinking, changed how questions were framed, introduced new methods, engendered important debate, or performed a heuristic function for other scholars.

A fifth characteristic of literature reviews concerns how a paper is *organized*. Reviews may be arranged historically, so that topics are introduced in the chronological order in which they appeared in the literature, conceptually, so that works relating to the same abstract ideas appear together, or methodologically, so that works employing similar methods are grouped together.

Finally, the intended audiences of reviews can vary. Reviews can be written for groups of specialized researchers, general researchers, policy makers, practitioners, or the general public. As reviewers move from addressing specialized researchers to addressing the general public, they employ less technical jargon and detail, while often paying greater attention to the implications of the work being covered.

1.2.2. A Definition of Research Synthesis

The terms *research synthesis*, *research review*, and *systematic review* are often used interchangeably in the social science literature, though they sometimes connote subtly different meanings. Regrettably, there is no consensus about whether these differences are really meaningful. Therefore, we will use the term *research synthesis* most frequently throughout this book. The reason for this choice is simple. In addition to its use in the context of research synthesis, the term *research review* is used as well to describe the activities of evaluating the quality of research. For example, a journal editor will obtain research reviews when deciding whether to publish a manuscript. Because research syntheses often include this type of evaluative review of research, using the term *research synthesis* avoids confusion. The term *systematic review* is less often used in the context of research evaluation, though the confusion is still there, and also it is a term less familiar to social scientists than is *research synthesis*.

A research synthesis can be defined as the conjunction of a particular set of literature review characteristics. Most definitional about research syntheses are their primary focus and goal: research syntheses attempt to integrate

empirical research for the purpose of creating generalizations. Implicit in this definition is the notion that seeking generalizations also involves seeking the limits of generalizations. Also, research syntheses almost always pay attention to relevant theories, critically analyze the research they cover, try to resolve conflicts in the literature, and attempt to identify central issues for future research. According to Derek Price, research syntheses are intended to “replace those papers that have been lost from sight behind the research front” (1965, 513). Research synthesis is one of a broad array of integrative activities that scientists engage in; as demonstrated in the quote leading this chapter, its intellectual heritage can be traced back at least as far as Aristotle.

Using the taxonomy described, we can make further specifications concerning the type of research syntheses that are the focus of this book. With regard to perspective, readers will note that much of the material is meant to help synthesists produce statements about evidence that are neutral in perspective, that is, less likely to be affected by bias or by their own subjective outlooks. The material on searching the literature for evidence is meant to help synthesists uncover all the evidence, not simply positive studies (that might be overrepresented in published research) or evidence that is easy for them to find (that might be overly sympathetic to their point of view or disciplinary perspective). The material on the reliability of extracting information from research reports and how methodological variations in research should be handled is meant to increase the interjudge reliability of these processes. The methods proposed for the statistical integration of findings is meant to ensure the same rules about data analysis are consistently applied to all conclusions about what is and is not significant, both statistically and practically. Finally, the material on explicit and exhaustive reporting of the methods used in syntheses is meant to assist users in evaluating if or where subjectivity may have crept into the process and to replicate findings for themselves if they choose to do so.

Finally, the term *meta-analysis* often is used as a synonym for research synthesis. However, in this volume, it will be used in its more precise and original meaning—to describe the quantitative procedures that a research synthesist may use to statistically combine the results of studies. Gene Glass coined the term to refer to “the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings” (1976, 3). The authors of the *Handbook* reserve the term to refer specifically to statistical analysis in research

synthesis and not to the entire enterprise of research synthesis. Not all research syntheses are appropriate for meta-analysis.

1.3. A BRIEF HISTORY OF RESEARCH SYNTHESIS AS A SCIENTIFIC ENTERPRISE

1.3.1 Early Developments

In 1971, Kenneth Feldman demonstrated remarkable prescience: “Systematically reviewing and integrating . . . the literature of a field may be considered a type of research in its own right—one using a characteristic set of research techniques and methods” (86). He described four steps in the synthesis process: sampling topics and studies, developing a scheme for indexing and coding material, integrating the studies, and writing the report.

The same year, Richard Light and Paul Smith (1971) presented what they called a cluster approach to research synthesis, meant to redress some of the deficiencies in the existing strategies for integration. They argued that if treated properly the variation in outcomes among related studies could be a valuable source of information—rather than merely a source of consternation as it appeared to be when treated with traditional cumulating methods.

Three years later, Thomas Taveggia struck a complementary theme:

A methodological principle overlooked by [synthesists] . . . is that research results are probabilistic. What this principle suggests is that, in and of themselves, the findings of any single research are meaningless—they may have occurred simply by chance. It also follows that, if a large enough number of researches has been done on a particular topic, chance alone dictates that studies will exist that report inconsistent and contradictory findings! Thus, what appears to be contradictory may simply be the positive and negative details of a distribution of findings. (1974, 397–98)

Taveggia went on to describe six common problems in research synthesis: selecting research; retrieving, indexing, and coding information from studies; analyzing the comparability of findings; accumulating comparable findings; analyzing distributions of results, and; reporting of results.

The development of meta-analytic techniques extends back further in time but their routine use by research synthesists is also relatively recent. Where Glass gave us the term meta-analysis in 1976, Ingram Olkin pointed out that ways to estimate effect sizes have existed since the turn of

the century (1990). For example, Pearson took the average of estimates from five separate samples of the correlation between inoculation for enteric fever and mortality (1904). He used this average to better estimate the typical effect of inoculation and to compare this effect with that of inoculation for other diseases. Early work on the methodology for combination of estimates across studies includes papers in the physical sciences by Raymond Birge (1932) and in statistics by William G. Cochran (1937), Frank Yates and Cochran (1938), and Cochran (1954).

Methods for combining probabilities across studies also have a long history, dating at least from procedures suggested in Leonard Tippett’s *The Method of Statistics* (1931) and Ronald Fisher’s *Statistical Methods for Research Workers* (1932). The most frequently used probability-combining method was described by Frederick Mosteller and Robert Bush in the first edition of the *Handbook of Social Psychology* more than fifty years ago (1954).

Still, the use of quantitative synthesis techniques in the social sciences was rare before the 1970s. Late in that decade, several applications of meta-analytic techniques captured the imagination of behavioral scientists: in clinical psychology, Mary Smith and Gene Glass’s meta-analysis of psychotherapy research (1977); in industrial-organizational psychology, Frank Schmidt and John Hunter’s validity generalization of employment tests; in social psychology, Robert Rosenthal and Donald Rubin’s integration of interpersonal expectancy effect research (1977), and, in education, Glass and Smith’s synthesis of the literature on class size and achievement (1978).

1.3.2 Research Synthesis Comes of Age

Two papers that appeared in the *Review of Educational Research* in the early 1980s brought the meta-analytic and synthesis-as-research perspectives together. The first proposed six synthesis tasks “analogous to those performed during primary research” (Jackson 1980, 441). Gregg Jackson portrayed meta-analysis as an aid to the task of analyzing primary studies but emphasized its limitations as well as its strengths. Also noteworthy about his paper was his use of a sample of thirty-six articles from prestigious social science periodicals to examine the methods used in integrative empirical syntheses. He reported, for example, that only one of the thirty-six syntheses reported the indexes or retrieval systems used to locate primary studies. He concluded that “relatively little

thought has been given to the methods for doing integrative reviews. Such reviews are critical to science and social policy making and yet most are done far less rigorously than is currently possible” (459).

Harris Cooper drew the analogy between research synthesis and primary research to its logical conclusion (1982). He presented a five stage model of synthesis as a research project. For each stage, he codified the research question asked, its primary function in the synthesis, and the procedural differences that might cause variation in synthesis conclusions. In addition, Cooper applied the notion of threats-to-inferential-validity—which Donald Campbell and Julian Stanley introduced for evaluating the utility of primary research designs—to the conduct of research synthesis (1966; also see Shadish, Cook, and Campbell 2002). Cooper identified ten threats to validity specifically associated with synthesis procedures that might undermine the trustworthiness of a research synthesis’ findings. He also suggested that other threats might exist and that the validity of any particular synthesis could be threatened by consistent deficiencies in the set of studies that formed its database. Table 1.2 presents a recent revision of this schema (Cooper 2007) that proposes a six stage model that separates the original analysis and interpretation stage into two distinct stages.

The first half of the 1980s also witnessed the appearance of four books primarily devoted to meta-analytic methods. In 1981, Gene Glass, Barry McGaw, and Mary Smith presented meta-analysis as a new application of analysis of variance and multiple regression procedures, with effect sizes treated as the dependent variable. In 1982, John Hunter, Frank Schmidt, and Greg Jackson introduced meta-analytic procedures that focused on comparing the observed variation in study outcomes to that expected by chance (the statistical realization of Taveggia’s point) and correcting observed effect size estimates and their variance for known sources of bias (for example, sampling error, range restrictions, unreliability of measurements). In 1984, Rosenthal presented a compendium of meta-analytic methods covering, among other topics, the combining of significance levels, effect size estimation, and the analysis of variation in effect sizes. Rosenthal’s procedures for testing moderators of variation in effect sizes were not based on traditional inferential statistics, but on a new set of techniques involving assumptions tailored specifically for the analysis of study outcomes. Finally, in 1985, with the publication of *Statistical Procedures for Meta-Analysis*, Hedges and Olkin helped to elevate the quantitative synthesis of research to

an independent specialty within the statistical sciences. This book, summarizing and expanding nearly a decade of programmatic developments by the authors, not only covered the widest array of meta-analytic procedures but also presented rigorous statistical proofs establishing their legitimacy.

Another text that appeared in 1984 also helped elevate research synthesis to a more rigorous level. Richard Light and David Pillemer focused on the use of research syntheses to help decision-making in the social policy domain. Their approach placed special emphasis on the importance of meshing both numbers and narrative for the effective interpretation and communication of synthesis results.

Several other books appeared on the topic of meta-analysis during the 1980s and early 1990s. Some of these treated the topic generally (for the most recent editions, see, for example, Cooper 2009; Hunter and Schmidt 2004; Lipsey and Wilson 2001; Wolf 1986), some treated it from the perspective of particular research design conceptualizations (for example, Mullen 1989), some were tied to particular software packages (for example, Johnson 1993), and some looked to the future of research synthesis as a scientific endeavor (for example, Wachter and Straf 1990; Cook et al. 1992). The first edition of the *Handbook of Research Synthesis* appeared in 1994.

Readers interested in a popular history of meta-analysis can consult Morton Hunt’s 1997 *How Science Takes Stock: The Story of Meta-Analysis*. In 2006, Mark Petticrew and Helen Roberts provided a brief “unsystematic history of systematic reviews” that includes discussion of early attempts to appraise the quality of studies and synthesize results. In 1990, Olkin did the same for meta-analysis.

Literally thousands of research syntheses have been published since the first edition of the *Handbook* in 1994. Figure 1.1 presents some evidence of the increasing impact of research syntheses on knowledge in the sciences and social sciences. The figure is based on entries in the Science Citation Index Expanded and the Social Sciences Citation Index, according to the Web of Science reference database (2006). It charts the growth in the number of citations to documents including the terms *research synthesis*, *systematic review*, *research review*, and *meta-analysis* in their title or abstract during the years 1995 to 2005 and this number of citations divided by the total number of documents in the two databases. The figure indicates that both the total number of citations and the number of citations per document in the databases has risen every year without exception. Clearly, the role that research syntheses play in our knowledge claims is large and growing larger.

Table 1.2 Research Synthesis Conceptualized as a Research Process

Stage	Stage Characteristics		
	Research Question	Primary Function	Procedural Variation
Define the Problem	What research evidence will be relevant to the problem or hypothesis of interest in the synthesis?	Define the variables and relationships of interest so that relevant and irrelevant studies can be distinguished	Variation in the conceptual breadth and detail of definitions might lead to differences in the research operations deemed relevant and/or tested as moderating influences
Collect the Research Evidence	What procedures should be used to find relevant research?	Identify sources (e.g., reference databases, journals) and terms used to search for relevant research and extract information from reports	Variation in searched sources and extraction procedures might lead to systematic differences in the retrieved research and what is known about each study
Evaluate the Correspondence between Methods and Implementation of Studies and the Desired Synthesis Inferences	What retrieved research should be included or excluded from the synthesis based on the suitability of the methods for studying the synthesis question or problems in research implementation?	Identify and apply criteria to separate correspondent from incommensurate research results	Variation in criteria for decisions about study inclusion might lead to systematic differences in which studies remain in the synthesis
Analyze (Integrate) the Evidence from Individual Studies	What procedures should be used to summarize and integrate the research results?	Identify and apply procedures for combining results across studies and testing for differences in results between studies	Variation in procedures used to analyze results of individual studies (narrative, vote count, averaged effect sizes) can lead to differences in cumulative results
Interpret the Cumulative Evidence	What conclusions can be drawn about the cumulative state of the research evidence?	Summarize the cumulative research evidence with regard to its strength, generality, and limitations	Variation in criteria for labeling results as important and attention to details of studies might lead to differences in interpretation of findings
Present the Synthesis Methods and Results	What information should be included in the report of the synthesis?	Identify and apply editorial guidelines and judgment to determine the aspects of methods and results readers of the synthesis report need to know	Variation in reporting might lead readers to place more or less trust in synthesis outcomes and influences others ability to replication results

SOURCE: Cooper 2007.

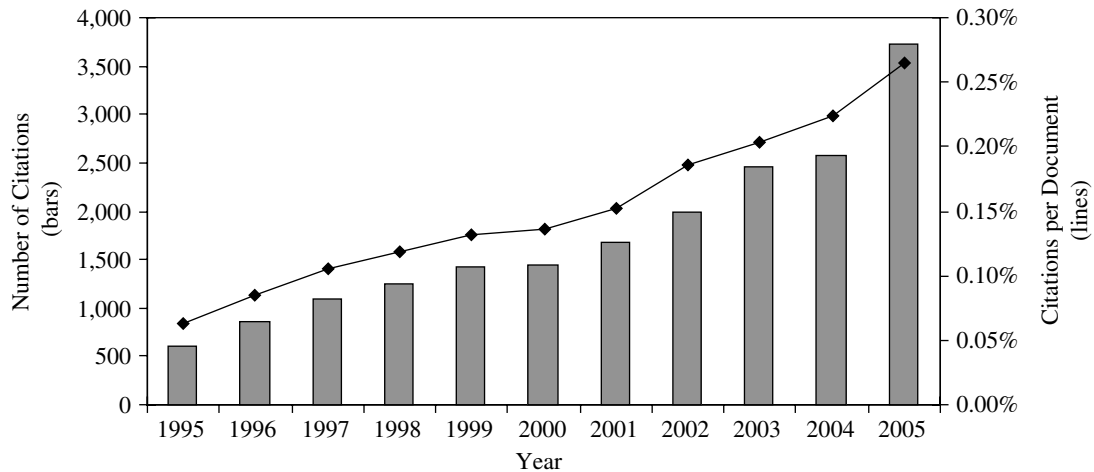


Figure 1.1 Citations to Articles, per Document

SOURCE: Authors' compilation.

NOTE. Based on entries in the Science Citation Index Expanded and the Social Sciences Citation Index, according to the Web of Science reference database (retrieved June 28, 2006). Bars chart the growth in the number of citations to documents including the terms *research synthesis*, *systematic review*, *research review* or *meta-analysis* in their title or abstract during the years following the publication of the first edition of the *Handbook of Research Synthesis*. Lines chart this number of citations divided by the total number of documents in the two databases.

In the past two decades the use of research synthesis has spread from psychology and education through many disciplines, especially the medical sciences and social policy analysis. Indeed, the development of scientific methods for research synthesis has its own largely independent history in the medical sciences (see Chalmers, Hedges, and Cooper 2002). A most notable event in medicine was the establishment of the U.K. Cochrane Centre in 1992. The Centre was meant to facilitate the creation of an international network to prepare and maintain systematic reviews of the effects of interventions across the spectrum of health care practices. At the end of 1993, an international network of individuals, called the Cochrane Collaboration (<http://www.cochrane.org/index.htm>), emerged from this initiative (Chalmers 1993; Bero and Rennie 1995). By 2006, the Cochrane Collaboration was an internationally renowned initiative with 11,000 people contributing to its work, in more than ninety countries. It is now the leading producer of research syntheses in health care and is considered by many to be the gold standard for determining the effectiveness of different health care interventions. Its library of systematic reviews numbers in the thousands. In 2000, an initiative called the Campbell Collaboration

(<http://www.campbellcollaboration.org/>) was begun with similar objectives for the domain of social policy analysis, focusing initially on policies concerning education, social welfare, and crime and justice.

Because of the efforts of scholars who chose to apply their skills to how research syntheses might be improved, syntheses written since the 1980s have been held to standards far more demanding than those applied to their predecessors. The process of elevating the rigor of syntheses is certain to continue into the twenty-first century.

1.3.3 Rationale for the *Handbook*

The Handbook of Research Synthesis and Meta-Analysis is meant to be the definitive volume for behavioral and social scientists intent on applying the synthesis craft. It distills the products of thirty years of developments in how research integrations should be conducted so as to minimize the chances of conclusions that are not truly reflective of the cumulated evidence. Research synthesis in the 1960s was at best an art, at worst a form of yellow journalism. Today, the summarization and integration of studies is viewed as a research process in its own right, is held to the standards of a scientific endeavor, and entails

the application of data gathering and analyses techniques developed for its unique purpose.

Numerous excellent texts on research synthesis exist. None, however, are as comprehensive as this volume. Some texts focus on statistical methods. These often emphasize different aspects of statistical integration (for example, combining probabilities, regression-analog models, estimating population effects from sampled effects with known biases) and often approach research accumulation from different perspectives. Although these texts are complete within their domains, no single sourcebook describes and integrates all the meta-analytic approaches.

This volume incorporates quantitative statistical techniques from all the synthesis traditions. It brings the leading authorities on the various meta-analytic perspectives together in a single volume. As such, it is an explicit statement by its authors that all the statistical approaches share a common assumptive base. The common base is not only statistical but also philosophical. Philosophically, all the approaches rest on the presupposition that research syntheses need to be held to the same standards of rigor, systematicity, and transparency as the research on which they are based. The second and later users of data must be held as accountable for the validity of their methods as were the first.

Several problems arising in the course of conducting a quantitative synthesis have not received adequate treatment in any existing text. These include nonindependence of data sets, synthesis of multivariate data sets, and sensitivity analysis, to name just a few. Every research synthesist faces these problems, and strategies have been developed for dealing with them. Some of their solutions are published in widely scattered journals, others are passed on to colleagues through informal contacts. They have never received complete treatment within the same text. The *Handbook* brings these topics together for the first time.

Further, texts focusing on the statistical aspects of integration tend to give only passing consideration to other activities of research synthesis. These activities include: the unique characteristics of problem formulation in research synthesis, methods of literature search, coding and evaluation of research reports, and the meaningful interpretation and effective communication of synthesis results. The existing texts that focus on these aspects of research synthesis tend not to be comprehensive in their coverage of statistical issues. Fully half of the chapters in this volume deal with issues that are not statistical, evidencing the authors' collective belief that high quality

syntheses require considerably more than just the application of quantitative procedures.

Finally, the handbook is meant for those who carry out research syntheses. Discussions of theory and proof are kept to a minimum in favor of descriptions of the practical mechanics needed to apply well the synthesis craft. The chapters present multiple approaches to problem solving and discuss the strengths and weaknesses of each approach. Readers with a comfortable background in analysis of variance and multiple regression and who have access to a research library should find the chapters accessible. The *Handbook* authors want to supply working synthesists with the needed expertise to interpret their blueprints, to wield their mortar hoe and trowel.

1.4 STAGES OF RESEARCH SYNTHESIS

The description of the stages of research synthesis presented in table 1.2 provides the conceptual organization of the handbook. In this section, we raise the principal issues associated with each stage. This allows us to briefly introduce the content of each of the chapters that follow.

1.4.1 Problem Formulation

Formulating a problem in research synthesis is constrained by one major factor: primary research on a topic must exist before a synthesis can be conducted. How much research? If the research question is important, it would be interesting to know how much research there is on the problem, even if the answer was none at all. The methods of meta-analysis can be applied to literatures containing as few as two hypothesis tests. Under certain circumstances—for instance, researchers synthesizing a pair of replicate studies from their own lab—the use of meta-analysis in this fashion might be legitimate. Yet, most scientists would argue that the benefits of such a synthesis would be limited (and its chances for publication even more limited).

A more general answer to the “how much research” question is that it varies depending on a number of characteristics of the problem. All else being equal, conceptually broad topics would seem to profit from a synthesis only after the accumulation of a more varied and larger number of studies than would narrowly defined topics. Similarly, literatures that contain diverse types of operations also would seem to require a relatively large number of studies before relatively firm conclusions could be drawn from a synthesis. Ultimately, the arbiter of whether

a synthesis is needed will not be numerical standards, but the fresh insights a synthesis can bring to a field. Indeed, although a meta-analysis cannot be performed without data, many social scientists see value in empty syntheses that point to important gaps in our knowledge. When done properly, empty syntheses should have proceeded through the stages of research synthesis, including careful problem formulation and a thorough literature search.

Once enough literature on a problem has collected, the challenge—and promise—of research synthesis becomes evident. The problems that constrain primary researchers—small and homogeneous samples, limited time and money for turning constructs of interest into operations—are less severe for synthesists. They can capitalize on the diversity in methods that has occurred naturally across primary studies. The heterogeneity of methods across studies may permit tests of theoretical hypotheses concerning the moderators and mediators of relations that have never been tested in any single primary study. Conclusions about the population and ecological validity of relations uncovered in primary research may also receive more thorough tests in syntheses.

Part II of the handbook focuses on issues in problem formulation. In chapter 2, Harris Cooper discusses in detail the issues mentioned above. In chapter 3, Larry Hedges looks at the implications of different problem definitions for how study results will be statistically modeled. The major issues involve the populations of people and measurements that are the target of a synthesis' inferences; how broadly the key constructs are defined, especially in terms of whether fixed or random effect models are envisioned; and how choices among models influence the precision of estimates and the statistical power of meta-analytic tests.

1.4.2. Literature Search

The literature search is the stage of research synthesis that is most different from primary research. Still, culling through the literature for relevant studies is not unlike gathering a sample of primary data. The target of a literature search that is part of a synthesis attempting exhaustive coverage would be all the research conducted on the topic of interest.

Cooper found that about half of the reviewers he surveyed claimed that they conducted their search with the intention of identifying all or most of the relevant literature (1987). About three-quarters said they did not stop searching until they felt they had accomplished their goal.

About one-third said they stopped their search when they felt their understanding and conclusions about the topic would not be affected by additional material. About one-sixth said they stopped searching when retrieval became unacceptably difficult.

In contrast to the relatively well-defined sampling frames available to primary researchers, literature searchers confront the fact that any single source of primary reports will lead them to only a fraction of the relevant studies, and a biased fraction at that. For example, among the most egalitarian sources of literature are the reference databases, such as PsycINFO, ERIC, and Medline. Still, these broad, non-evaluative systems exclude much of the unpublished literature. Conversely, the least equitable literature searching technique involves accessing one's close colleagues and other researchers with an active interest in the topic area. In spite of the obvious biases, there is no better source of unpublished and recent works. Further complicating the sampling frame problem for synthesist is the fact that the relative utility and biases associated with any single source will vary as a function of characteristics of the research problem, including for example how long the topic has been the focus of study and whether the topic is interdisciplinary in nature.

These problems imply that research synthesists must carefully consider multiple channels for accessing literature and how the channels they choose complement one another. The three chapters in part III are devoted to helping the synthesist consider and carry out this unique task. In chapter 4, Howard White presents an overview of searching issues from the viewpoint of an information scientist. Both chapter 5, by Jeffrey Reed and Pam Baxter, and chapter 6, by Hannah Rothstein and Sally Hopewell, present in detail the practical considerations of how to scour our brickyards and quarries.

1.4.3 Data Evaluation

Once the synthesists have gathered the relevant literature, they must extract from each document those pieces of information that will help answer the questions that impel research in the field. The problems faced during data coding provide a strong test of the synthesist's competence, thoughtfulness, and ingenuity. The solutions found for these problems will have considerable influence on the contribution of the synthesis. Part IV contains four chapters on data coding and evaluation.

The aspect of coding studies that engenders the most debate involves how synthesists should represent

differences in the design and implementation of primary studies. What is meant by quality in evaluating research methods? Should studies be weighted differently if they differ in quality? Should studies be excluded if they contain too many flaws? How does one rate the quality of studies described in incomplete research reports? In chapter 7, Jeff Valentine examines the alternative approaches available to synthesisists for representing primary research methodology. Beyond quality judgments, synthesisists make decisions about the classes of variables that are of potential interest to them. These can relate to variables that predict outcomes, potential mediators of effects, and the differences in how outcomes are conceptualized (and, therefore, measured). If a synthesisist chooses not to code a particular feature of studies, then it cannot be considered in the analysis of results. General guidelines for what information should be extracted from primary research reports are difficult to develop, beyond those that are very abstract. Instead, direction will come from the issues that arise in the particular literature, coupled with the synthesisist's personal insights into the topic. Still, there are commonalities that emerge in how these decisions are forged. Mark Lipsey, in chapter 8 ("Identifying Variables and Analysis Opportunities"), and David Wilson, in chapter 9 ("Systematic Coding"), present complementing templates for what generally should be included in coding frames.

Once decisions on what to code have been made, synthesisists need to consider how to carry out the coding of the literature and how to assess the trustworthiness with which the coding frame is implemented. There are numerous indexes of coder reliability available, each with different strengths and weaknesses. Chapter 10, by Robert Orwin and Jack Vevea, describes strategies for reducing the amount of error that enters a synthesis during the coding of the literature's features. Orwin and Vevea's description of reliability assessment focuses on three major approaches: measuring interrater agreement; obtaining confidence ratings from coders; and conducting sensitivity analysis, or using multiple coding strategies to determine whether a synthesis' conclusions are robust across different coding decision rules.

1.4.4 Data Analysis

As our brief history of research synthesis revealed, the analysis of accumulated research outcomes has been a separate area of specialization within statistics. Three decades ago, the actual mechanics of integrating research

usually involved intuitive processes taking place inside the heads of the synthesisists. Meta-analysis made these processes public and based them on explicit, shared, statistical assumptions (however well these assumptions were met). We would not accept as valid a primary researcher's conclusion if it were substantiated solely by the statement "I looked at the treatment and control scores and I think the treated group did better." We would demand some sort of statistical test (for example, a *t*-test) to back up the claim. Likewise, we no longer accept "I examined the study outcomes and I think the treatment is effective" as sufficient warrant for the conclusion of a research synthesis.

Part V contains chapters that cover the components of synthesis dealing with combining study results. Brad Bushman and Morgan Wang, in chapter 11, describe procedures based on viewing the outcomes of studies dichotomously, as either supporting or refuting a hypothesis. The next two chapters cover methods for estimating the magnitude of an effect, commonly called the effect size. Cohen defined an effect size as "the *degree* to which the phenomenon is present in the population, or the degree to which the null hypothesis is false" (1988, 9–10). In chapter 12, Michael Borenstein describes the parametric measures of effect size. In chapter 13, Joseph Fleiss and Jesse Berlin cover the measures of effect size for categorical data, paying special attention to the estimate of effect based on rates and ratios.

To most research synthesisists, the search for influences on study results is the most exciting and rewarding part of the process. The four chapters of part VI all deal with techniques for analyzing whether and why there are differences in the outcomes of studies. As an analog to analysis of variance or multiple regression procedures, effect sizes can be viewed as dependent (or criterion) variables and the features of study designs as independent (or predictor) variables. Because effect-size estimates do not all have the same sampling uncertainty, however, they cannot be combined using traditional inferential statistics. In chapter 14, William Shadish and Keith Haddock describe procedures for averaging effects and generating confidence intervals around these population estimates. In chapter 15, Spyros Konstantopoulos and Larry Hedges discuss analysis strategies when fixed-effects models are used. In chapter 16, Stephen Raudenbush describes the strategies appropriate for random-effects models.

Further, effect-size estimates may be affected by factors that attenuate their magnitudes. These may include, for example, a lack of reliability in the measurement

instruments or restrictions in the range of values in the subject sample. Attenuating biases may be estimated and corrected using the procedures that Frank Schmidt, Huy Le, and In-Sue Oh describe in chapter 17.

Part VII includes three chapters that focus on some special issues in the statistical treatment of research outcomes. In chapter 18, Larry Hedges discusses combining special types of studies, in particular the increasing use of cluster- or group-randomized trials in social research and the implication of this for the definition of effect sizes, their estimation, and particularly the computation of standard errors of the estimates. Chapter 19 tackles the vexing problem of what to do when the same study (or laboratory, or researcher, if one wishes to push the point) provides more than one effect size to a meta-analysis. Because estimates from the same study may not be independent sources of information about the underlying population value, their contributions to any overall estimates must be adjusted accordingly. In chapter 19, Leon Gleser and Ingram Olkin describe both simple and more complex strategies for dealing with dependencies among effect sizes.

Sometimes, the focus of a research synthesis will not be a simple bivariate relation but a model involving numerous variables and their interrelations. The model could require analysis of more than one simple relation and/or complex partial or second-order relations. In chapter 20, Betsy Becker looks at what models in meta-analysis are, why they should be used, and how they affect meta-analytic data analysis.

1.4.5 Interpretation of Results

Estimating and averaging effect sizes and searching for moderators of their variability is how the interpretation of cumulative study results begins. However, it must be followed by other procedures that help the synthesists properly interpret their outcomes. Proper interpretation of the results of a research synthesis requires careful use of declarative statements regarding claims about the evidence, specification of what results warrants each claim, and any appropriate qualifications to claims that need to be made. In part VIII, three important issues in data interpretation are examined.

Even the most trustworthy coding frame cannot solve the problem of missing data. Missing data will arise in every research synthesis. Reporting of study features and results, as now practiced, leaves much to be desired. Although the emergence of research synthesis has led to

many salutary changes, most primary researchers still do not think about a research report in terms of the needs of the next user of the data. Also, problems with missing data would still occur even if the needs of the synthesist were the paramount consideration in constructing a report. There is simply no way primary researchers can anticipate the subtle nuances in their methods and results that might be identified as critical to piecing together the literature five, ten, or more years later. In chapter 21, Theresa Piggott describes how lost information can bear on the results and interpretation of a research synthesis. She shows how the impact of missing data can be assessed by paying attention to some of its discoverable characteristics. She also describes both simple and complex strategies for handling missing data in research synthesis. Joel Greenhouse and Satish Iyengar, in chapter 22, describe sensitivity analysis in detail, including several diagnostic procedures to assess the sensitivity and robustness of the conclusions of a meta-analysis. Alex Sutton, in chapter 23, examines publication bias and describes procedures for identifying whether data censoring has had an impact on the meta-analytic result. He also details some imputation procedures that can help synthesists estimate what summary effect size estimates might have been in the absence of the selective inclusion of studies.

The chapters in part IX bring us full circle, returning us to the nature of the problem that led to undertaking the research synthesis in the first place. Collectively, these chapters attempt to answer the question as to what can be learned from research synthesis. In chapter 24, Wendy Wood and Alice Eagly examine how theories can be tested in research synthesis. In chapter 25, David Cordray and Paul Morphy look at research synthesis for policy analysis. These chapters are meant to help synthesists think about the substantive questions for which the readers of their syntheses will be seeking answers.

1.4.6 Public Presentation

Presenting the background, methods, results, and meaning of a research synthesis' findings provide the final challenges to the synthesists' skill and intellect. In chapter 26, the first of part X, Geoffrey Borman and Jeffrey Grigg address the visual presentation of research synthesis results. Charts, graphs, and tables, they explain, should be used to summarize the numbers in a meta-analysis, along with a careful intertwining of narrative explication to contextualize the summaries. Proper use of interpretive devices can make a complex set of results easily accessible

to the intelligent reader. Likewise, inadequate use of these aides can render incomprehensible even the simplest conclusions. Chapter 27 describes the reporting format for research synthesis. As with the coding frame, there is no simple reporting scheme that fits all syntheses. However, certain commonalities do exist and Mike Clarke discusses them. Not too surprisingly, the organization that emerges bears considerable resemblance to that of a primary research report although, also obviously, the content differs dramatically.

1.5 CHAPTERS IN PERSPECTIVE

In part XI, the final two chapters of the handbook are meant to bring the detailed drawings of the preceding chapters onto a single blueprint. In chapter 28, Georg Matt and Thomas Cook present an expanded analysis of threats to the validity of research syntheses, provide an overall appraisal of how inferences from syntheses may be restricted or faulty, and bring together many of the concerns expressed throughout the book.

Finally, chapter 29 also attempts to draw this material together. In it, Harris Cooper and Larry Hedges look at some of the potentials and limitations of research synthesis. Special attention is paid to possible future developments in synthesis methodology, the feasibility and expense associated with conducting a sound research synthesis, and a broad-based definition of what makes a research synthesis good or bad.

No secret will be revealed by stating our conclusion in advance. If procedures for the synthesis of research are held to standards of objectivity, systematicity, and rigor, then our knowledge edifice will be made of bricks and mortar. If not, it will be a house of cards.

1.6 REFERENCES

- Bero, Lisa, and Drummond Rennie. 1995. "The Cochrane Collaboration: Preparing, Maintaining, and Disseminating Systematic Reviews of the Effects of Health Care." *Journal of the American Medical Association* 274(24): 1935–8.
- Birge, Raymond T. 1932. "The Calculation of Error by the Method of Least Squares." *Physical Review* 40(2): 207–27.
- Campbell, Donald T., and Julian C. Stanley. 1966. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- Chalmers, Iain. 1993. "The Cochrane Collaboration: Preparing, Maintaining and Disseminating Systematic Reviews of the Effects of Health Care." *Annals of the New York Academy of Sciences* 703: 156–63.
- Chalmers, Iain, Larry V. Hedges, and Harris Cooper. 2002. "A Brief History of Research Synthesis." *Evaluation and the Health Professions* 25(1): 12–37.
- Cochran, William G. 1937. "Problems Arising in the Analysis of a Series of Similar Experiments." *Journal of the Royal Statistical Society* 4(Supplement): 102–18.
- . 1954. "The Combination of Estimates from Different Experiments." *Biometrics* 10(1, March): 101–29.
- Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Cook, Thomas D., Harris M. Cooper, David S. Cordray, Heidi Hartmann, Larry V. Hedges, Richard J. Light, Thomas Louis, and Frederick Mosteller. 1992. *Meta-Analysis for Explanation: A Casebook*. New York: Russell Sage Foundation.
- Cooper, Harris M. 1982. "Scientific Guidelines for Conducting Integrative Research Reviews." *Review of Educational Research* 52(2, Summer): 291–302.
- . 1987. "Literature Searching Strategies of Integrative Research Reviewers: A First Survey." *Knowledge: Creation, Diffusion, Utilization* 8: 372–83.
- . 1988. "Organizing Knowledge Synthesis: A Taxonomy of Literature Reviews." *Knowledge in Society* 1: 104–26.
- . 2009. *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*. Thousand Oaks, Calif.: Sage Publications.
- . 2003. "Editorial." *Psychological Bulletin* 129(1): 3–9.
- . 2007. *Evaluating and Interpreting Research Syntheses in Adult Learning and Literacy*. Boston, Mass.: National College Transition Network, New England Literacy Resource Center/World Education, Inc.
- Feldman, Kenneth A. 1971. "Using the Work of Others: Some Observations on Reviewing and Integrating." *Sociology of Education* 4(1, Winter): 86–102.
- Fisher, Ronald A. 1932. *Statistical Methods for Research Workers*, 4th ed. London: Oliver & Boyd.
- Glass, Gene V. 1976. "Primary, Secondary, and Meta-Analysis." *Educational Researcher* 5(10): 3–8.
- Glass, Gene V., and Mary L. Smith. 1978. *Meta-Analysis of Research on the Relationship of Class Size and Achievement*. San Francisco: Far West Laboratory for Educational Research and Development.
- Glass, Gene V., Barry McGaw, and Mary L. Smith. 1981. *Meta-Analysis in Social Research*. Beverly Hills, Calif.: Sage Publications.

- Hedges, Larry V. 1987. "How Hard Is Hard Science, How Soft is Soft Science?" *American Psychologist* 42(5): 443–55.
- Hedges, Larry V., and Ingram Olkin. 1985. *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.
- Hunt, Morton. 1997. *How Science Takes Stock: The Story of Meta-Analysis*. New York: Russell Sage Foundation.
- Hunter, John E., and Frank L. Schmidt. 2004. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, 2nd ed. Thousand Oaks, Calif.: Sage Publications.
- Hunter, John E., Frank L. Schmidt, and Greg B. Jackson. 1982. *Meta-Analysis: Cumulating Research Findings Across Studies*. Beverly Hills, Calif.: Sage Publications.
- Jackson, Gregg B. 1980. "Methods for Integrative Reviews." *Review of Educational Research* 50: 438–60.
- Johnson, Blair T. 1993. *DSTAT: Software for the Meta-Analytic Review of Research*. Hillsdale, N.J.: Lawrence Erlbaum.
- Light, Richard J., and David B. Pillemer. 1984. *Summing Up: The Science of Reviewing Research*. Cambridge, Mass.: Harvard University.
- Light, Richard J., and Paul V. Smith. 1971. "Accumulating Evidence: Procedures for Resolving Contradictions Among Research Studies." *Harvard Educational Review* 41(4): 429–71.
- Lipsey, Mark W., and David B. Wilson. 2001. *Practical Meta-Analysis*. Thousand Oaks, Calif.: Sage Publications.
- Manten, A. A. 1973. "Scientific Literature Review." *Scholarly Publishing* 5: 75–89.
- Mosteller, Frederick, and Robert Bush. 1954. "Selected Quantitative Techniques." In *Handbook of Social Psychology*, vol. 1, *Theory and Method*, edited by Gardner Lindzey. Cambridge, Mass.: Addison-Wesley.
- Mullen, Brian. 1989. *Advanced BASIC Meta-Analysis*. Hillsdale, N.J.: Lawrence Erlbaum.
- Olkin, Ingram. 1990. "History and Goals." In *The Future of Meta-Analysis*, edited by Kenneth W. Wachter and Miron L. Straf. New York: Russell Sage Foundation.
- Pearson, Karl. 1904. "Report on Certain Enteric Fever Inoculation Statistics." *British Medical Journal* 3: 1243–6.
- Petticrew, Mark, and Helen Roberts. 2006. *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford: Blackwell Publishing.
- Price, Derek J. de Solla. 1965. "Networks of Scientific Papers." *Science* 149(3683): 56–64.
- Rosenthal, Robert. 1984. *Meta-Analytic Procedures for Social Research*. Beverly Hills, Calif.: Sage Publications.
- Rosenthal, Robert, and Donald B. Rubin. 1978. "Interpersonal Expectancy Effects: The First 345 Studies." *The Behavioral and Brain Sciences* 3: 377–86.
- Schmidt, Frank L., and John E. Hunter. 1977. "Development of a General Solution to the Problem of Validity Generalization." *Journal of Applied Psychology* 62: 529–40.
- Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, Mass.: Houghton Mifflin.
- Smith, Mary L., and Gene V. Glass. 1977. "Meta-Analysis of Psychotherapy Outcome Studies." *American Psychologist* 32(9): 752–60.
- Taveggia, Thomas C. 1974. "Resolving Research Controversy Through Empirical Cumulation: Toward Reliable Sociological Knowledge." *Sociological Methods & Research* 2(4): 395–407.
- Tippett, Leonard Henry Caleb. 1931. *The Methods of Statistics*. London: Williams & Norgate.
- Wachter, Kenneth W., and Miron L. Straf, eds. 1990. *The Future of Meta-Analysis*. New York: Russell Sage Foundation.
- Wolf, Frederic M. 1986. *Meta-Analysis: Quantitative Methods for Research Synthesis*. Beverly Hills, Calif.: Sage Publications.
- Yates, Frank, and William G. Cochran. 1938. "The Analysis of Groups of Experiments." *Journal of Agricultural Science* 28: 556–80.