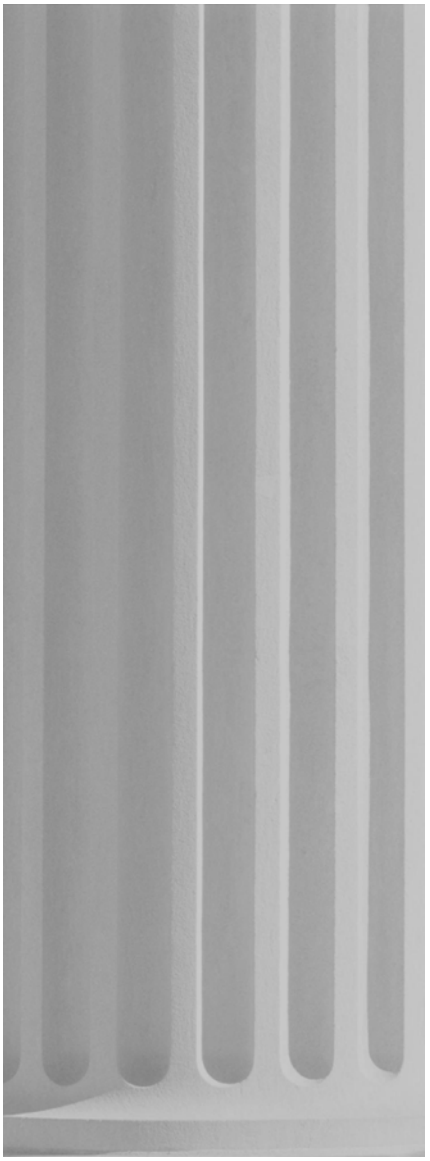
CHAPTER 16

Evaluation, Research Synthesis, Meta-analysis, and Evidence-Based Practice

KEY TERMS

evidence-based practice
meta-analysis
research synthesis
effect size
meta-analysis
secondary analysis
cost-effectiveness and cost-benefit analysis
impact evaluation
outcome evaluations
process evaluation
implementation evaluation
structured conceptualization
evaluability assessment
needs assessment
summative evaluations
formative evaluation
evaluation
evolutionary epistemology



OUTLINE

16-1 Background and Context

This book was primarily written to help you navigate the road to research and perhaps feel that you were in the company of a trusted guide along the way. Most of the stops on the research road have been places where you would meet ideas about such things as measurement, design, and analysis. I tried to pay some attention to where we'd been and where we were going in linking the chapters together, but mainly our view has been up-close to the matter at hand, topic by topic, study by study, chapter by chapter, step by step. But, then what? How can we use our research skills to address real-world problems and generate practically useful results that affect decision-making and policy development? We consider such applied uses of research methods in the section in this chapter on *evaluation*. Furthermore, imagine that the research road you've been on has come to a surprising turn: You can now see the entire journey for your research project, and you can see the projects and paths that others have taken in their research projects, too. Imagine the view as from a mountaintop, where you can see all of the terrain and take it in as a whole. Wouldn't that be something? Well, in this chapter we're also going to introduce you to this broader perspective. *Research synthesis*, *meta-analysis*, and *evidence-based practice* are all terms that reflect this mountaintop view.

This chapter addresses some of the most important innovations in contemporary research methods over the past decade, although most have histories that stretch way back to the late 1800s. Evaluation research is a subfield of social science that not only addresses the evidence-based practice question of "what works," but also has a long-standing focus on how things work and how well. Research synthesis is a systematic approach used to integrate the results of multiple studies, including qualitative and mixed methods approaches. Meta-analysis is a specific type of research synthesis that can be used when quantitative results are available from multiple studies. These methods are part of a new movement of "evidence-based practice" that encourages that practice in the form of research and that we make more effective use of research in practice. This movement holds the promise of a new era in the way our society integrates research into the way we live.

16-2 An Evolutionary, Ecological, Systems View of Research

This volume so far has emphasized how we would conduct a single research study. But what difference does a single study make? How do we use the results of a study

to advance knowledge? It should be clear by now that no study by itself can ever be considered perfect. Every study will have some threats to validity that cannot be completely addressed. So, how do we build our understanding and knowledge, a firm evidence base, from an imperfect, fallible study? The short answer to this question is that we don't. No single study is likely to resolve the complexities of a topic of interest. And that's the key to the answer—we don't rely on a single study to establish knowledge. Ultimately, we look at multiple studies. Research synthesis of multiple studies, meta-analysis of quantitative results as part of research synthesis, the dissemination and of the resulting evidence-based interventions, and the evaluation of their effects, are all essential to the building of our knowledge in science and in applied social research.

At the root of how a single study fits into the larger endeavor of human knowledge acquisition is a philosophical framework, a perspective about what we are trying to accomplish in research, a view that is based on the ideas of evolution and ecological systems thinking. Let's start with a basic idea, that of evolution. Just as species evolve, philosophers also have come to believe that ideas evolve. This is a notion called **evolutionary epistemology** (Campbell, 1974, 1988; Popper, 1985), the idea that our knowledge about the world evolves according to the same principles that describe the evolution of all life. In evolutionary theory generally we believe that there is variation among organisms. If the organism survives its environment, its unique variations have a higher chance of being preserved and passed on to others. If not, they become extinguished along with the organism. A similar process occurs with ideas. New ideas are suggested that are invariably based in some way on previous thinking. Sometimes they take hold. Other times they are criticized and rejected, or simply ignored altogether. There is, in effect, an evolutionary process of natural selection for ideas just as there is for species. A research project is simply the examination of one or more ideas (the questions or hypotheses). The results of our research might be accepted, be criticized and rejected, or ignored altogether. Biologists will tell you that evolution operates on populations of species. In each generation of a species there will be variation and some of this variation may be selected for and survive, and be promulgated through the species more broadly over time. The same thing might be said for research. For any given issue or problem we may over time generate a population of studies, a collection of studies that all differ slightly from each other (at the very least they are done with different people, in different settings and at different times). We want to look at these studies and try to decide which variations or studies are telling us something valuable, which should "survive" to contribute to our knowledge about the issue.

evolutionary epistemology
A philosophical position that holds that all knowledge evolves through evolutionary processes of natural selection.

Research synthesis and its quantitative sub-field of meta-analysis are designed to look at populations of studies to determine what they say as a species or group. In essence, research methods contribute studies that form this population and research synthesis and meta-analysis are critically important tools for the selection of survivor ideas in the evolution of our knowledge. Evidence-Based Practice is the process of disseminating or encouraging the use of the surviving knowledge or evidence from synthesis and meta-analysis. Evaluation can be viewed as the process of assessing or providing feedback about the effects of using programs that we established as potentially applicable in previous research. In addition, multiple evaluations on similar programs contribute to the ecology and provide more variation for us to use in understanding what works and evolving better knowledge and solutions to problems.

When we situate research as an evolutionary endeavor, and a research study as one of a population of similar studies that accrue over time to address some central problem or question, we are viewing a research project as part of an ecology of research, one element in a system of how people come to understand the world around us. This leads us to think a little differently about a specific research project. Increasingly, we are recognizing how important it is to see our research within an ecological and systems framework and to apply systems thinking principles and perspectives to our stand-alone research studies.

16-3 A Systems Perspective on Research

Trochim and colleagues (2006) recently defined *systems thinking* as "a general conceptual orientation concerned with the interrelationships between parts and their relationships to a functioning whole, often understood within the context of an even greater whole" (p. 539). This definition allows us to consider a range of systems, including large systems like the health-care system of our country, to personal systems such as our families, to biological systems that form the basis of life. From this point of view, research may be thought of as a systematic process embedded in many levels of dynamic interacting systems. As with most systems, there is some identifiable output, from the single study to the set of studies synthesized in the research base that may become practice guidelines. One of the ongoing discussions related to research synthesis is how to evaluate the quality of the input, that is, the quality of the primary study that produced the data to be integrated into the synthesis, and how the strengths and weaknesses of the input influenced the output.

Consider your research for a moment. You may be writing the justification for your study with regard to the impact your study may have on a system that provides care for people, like social, mental health, or medical care systems. You might have chosen this topic because it appears that the system is flawed or that the system that has produced the prior studies you have reviewed is deficient. But your work is embedded in many other system levels, from the systems of constructs or concepts that make up the programs and outcomes in your studies, to your university or college department, to the internal review board that will consider the ethical and perhaps scientific quality of your study, to a funding system that will hopefully support your study, to the systems that you must access to obtain your sample, and on to the systems that represent the potential avenues for dissemination of your results. Your study will be part of a larger "flow" within the research system in your field. Research synthesis is concerned with establishing a valid view of the overall flow of results on specific issues, whether those results are based on quantitative, qualitative, or mixed methods studies, and feeding the syntheses back to all of the parts of the system so they can better understand what is going on and make use of our evolving knowledge. Evidence-based practice is the translation of that flow of knowledge into ways of helping people with the practical problems of living.

16-4 The Link between Research and Practice

The link between research and practice is the most practical justification for research synthesis. In the hourglass model presented in the beginning of this book, the last step is to generalize back to the original problem. Research synthesis creates the possibility that studies can come full circle, and address the big questions that typically motivate the very specific studies that primary researchers do. Among all of the big questions, one central one expresses the link best: "What works?"

The importance of the science-practice link was expressed by the person who founded the very first psychological clinic in 1896, Lightner Witmer. Witmer noted that the progress of any basic science was integrally related to the success of the corresponding applied science and that the contributions of any science (basic or applied) would ultimately be judged by the contribution made to the human condition (Hayes, Barlow, & Nelson-Gray, 1999). In addition to medicine, nearly all of the social sciences and education have adopted some version of an evidence-based perspective. In fact, in applied psychology, the dominant training model has been called the "scientist-practitioner" model for more than 50 years (Hayes, Barlow, & Nelson-Gray, 1999).

16-5 History

In this section, I will review some of the additional background that has led us to our current methods of research synthesis. In addition to general acceptance of the value of the science-practice link, four important influences have come together to produce the advances in research synthesis methods:

1. The development of computing power, the Internet, and bibliographic databases.
2. Recognition of limits in traditional null hypothesis significance testing and overemphasis on significance levels.
3. Internal critics such as Hans Eysenck who questioned the value of psychotherapy and external pressures from government and other research funding agencies to be accountable.
4. Innovative researchers who have recognized the need for good answers to big questions and contributed methods. These include Smith and Glass (1977), Rosenthal and Rubin (1978), Schmidt and Hunter (1977), Light and Pillemer (1984), Cooper (1998), Lipsey and Wilson (2001), Kline (2004) Thompson (2006), and Stake (2006), among many others. They have contributed methods to facilitate synthesis and make it a lively and mainstream scientific activity.

We begin where many advances in scientific methods have been made, in agriculture, and then proceed to briefly review other fields where research synthesis and evidence-based practice have emerged.

16-6 Agriculture and Extension

Since its inception nearly a century ago, the United States Department of Agriculture (USDA) has managed a diverse set of research, education and community activities central to the lives of farmers and consumers through the Cooperative Extension service (http://www.csrees.usda.gov/index.html). Among the central goals stated in the original USDA legislation in 1914 was a statement that we would now recognize as part of the language of evidence-based practice: that the work of the Extension "... shall consist of the development of practical applications of research knowledge ..." (Dunifon, Duttweiler, Pillemer, Tobias, & Trochim, W. M. K., 2004, retrieved September 4, 2006 from http://www.joe.org/joe/2004april/a2.shtml). Currently, the focus of evidence-based practice Extension efforts includes traditional topics related to agriculture, but also more recently emerging issues such as food biosecurity (prevention of terrorist attacks on the nation's food supply). Thus, the view from agricultural extension suggests that we can agree with Norcross, Beutler, and Levant (2006), who likened evidence-based practice to clinical psychology, a field with "a long past but a short history" (p. 4).

16-7 Evaluation Research

Evaluation not only asks the evidence-based practice question of "what works," but also has a long history of developing methods of understanding how things work, and how well in practical situations and for real-world problems. It is a methodological area that is closely related to, but distinguishable from, more traditional social research. Evaluation uses many of the same methodologies used in traditional social research, but because evaluation takes place within a political and organizational context, it requires group skills, management ability, political dexterity, sensitivity to multiple stakeholders, and other skills that social research in general does not rely on as much. In the following sections, I introduce the idea of evaluation and some of the major terms and issues in the field.

evaluation
The systematic acquisition and assessment of information to provide useful feedback about some object.

16-7a Definitions of Evaluation

Probably the most frequently given definition of *evaluation* is "the systematic assessment of the worth or merit of some object." This definition is hardly perfect. Many types of evaluations do not *necessarily* result in an assessment of worth or merit—descriptive studies, implementation analyses, and formative evaluations, to name a few. Better perhaps is a definition that emphasizes the information-processing and feedback functions of evaluation. For instance, one might say that *evaluation* is "the systematic acquisition and assessment of information to provide useful feedback about some object."

Both definitions agree that evaluation is a systematic endeavor and both use the deliberately ambiguous term *object*, which could refer to a program, policy, technology, person, need, activity, and so on. The latter definition emphasizes acquiring and assessing information rather than assessing worth or merit because all evaluation work involves collecting and sifting through data, making judgments about the validity of the information and of inferences derived from it, whether or not an assessment of worth or merit results.

16-7b The Goals of Evaluation

The generic goal of most evaluations is to provide useful feedback to a variety of audiences, including sponsors, donors, client groups, administrators, staff, and other relevant constituencies. Most often, feedback is perceived as useful if it aids in decision making, but the relationship between an evaluation and its impact is not a simple one. Studies that seem critical sometimes fail to influence short-term decisions, and studies that initially seem to have no influence can have a delayed impact when more congenial conditions arise. Despite this, the broad consensus is that a major goal of evaluation should be to influence decision making or policy formulation through the provision of empirically driven feedback.

formative evaluation
Evaluations that strengthen or improve the object being evaluated. Formative evaluations are used to improve programs while they are still under development.

summative evaluations
Evaluations that examine the effects or outcomes of some program or treatment.

needs assessment
A structured empirical process for assessing the needs for a program or intervention.

evaluability assessment
A structured empirical process of assessing whether a formal evaluation is feasible in any specific context.

structured conceptualization
Any structured empirical method that can be used to develop a conceptual model or theory, often as the basis for a research project or evaluation.

implementation evaluation
A structured empirical process that assesses the degree to which a program or technology is implemented faithfully according to its conceptual specifications.

process evaluation
A structured empirical assessment of the process of delivering a program or technology.

16-7c Types of Evaluation

There are many different types of evaluations depending on the object being evaluated and the purpose of the evaluation. Perhaps the most important basic distinction in evaluation types is that between formative and summative evaluation. **Formative evaluations** strengthen or improve the object being evaluated; they help form it by examining the delivery of the program or technology, the quality of its implementation, and the assessment of the organizational context, personnel, procedures, inputs, and so on. **Summative evaluations**, in contrast, examine the effects or outcomes of some object. They summarize it by describing what happens subsequent to delivery of the program or technology, assessing whether the object can be said to have caused the outcome, determining the overall impact of the causal factor beyond only the immediate target outcomes, and estimating the relative costs associated with the object.

Formative evaluation includes several evaluation types:

- **Needs assessment** determines who needs the program, how great the need is, and what might work to meet the need.
- **Evaluability assessment** determines whether an evaluation is feasible and how stakeholders can help shape its usefulness.
- **Structured conceptualization** (for example, concept mapping) helps stakeholders define the program or technology, the target population, and the possible outcomes.
- **Implementation evaluation** monitors the fidelity of the program or technology delivery.
- **Process evaluation** investigates the process of delivering the program or technology, including alternative delivery procedures.

Summative evaluation can also be subdivided:

- **Outcome evaluations** investigate whether the program or technology caused demonstrable effects on specifically defined target outcomes.
- **Impact evaluation** is broader and assesses the overall or net effects—intended or unintended—of the program or technology as a whole.
- **Cost-effectiveness and cost-benefit analysis** address questions of efficiency by standardizing outcomes in terms of their dollar costs and values.
- **Secondary analysis** reexamines existing data to address new questions or use methods not previously employed.
- **Meta-analysis** integrates the outcome estimates from multiple studies to arrive at an overall or summary judgment on an evaluation question.

outcome evaluations
The formal assessment of whether a program or treatment caused the observed outcomes.

impact evaluation
The formal empirical assessment of the overall effects (intended and unintended) of a program or technology.

cost-effectiveness and cost-benefit analysis
Economic assessments of the costs of achieving specific outcomes and of these costs weighed against the estimated benefits.

secondary analysis
The formal empirical re-analysis of existing datasets to address new questions, replicate prior results or explore new analysis methods.

meta-analysis
A structured empirical analysis that integrates and summarizes results of multiple research studies or evaluations.

16-7d Evaluation Questions and Methods

Evaluators ask many different kinds of questions and use a variety of methods to address them. These are considered within the framework of formative and summative evaluation as presented in the previous section. In formative research the major questions and methodologies are as follows:

- *What is the definition and scope of the problem or issue, or what's the question?* Formulating and conceptualizing methods might be used, including brainstorming, focus groups, nominal group techniques, Delphi methods, brainwriting, stakeholder analysis, synectics, lateral thinking, input-output analysis, and concept mapping.
- *Where is the problem and how big or serious is it?* The most common method used here is needs assessment, which can include analysis of existing data sources and the use of sample surveys, interviews of constituent populations, qualitative research, expert testimony, and focus groups.
- *How should the program or technology be delivered to address the problem?* Some of the methods already listed apply here, as do detailing methodologies such as simulation techniques, or multivariate methods like multi-attribute utility theory or exploratory causal modeling; decision-making methods; and project planning and implementation methods such as flowcharting, PERT/CPM, and project scheduling.
- *How well is the program or technology delivered?* Qualitative and quantitative monitoring techniques, the use of management information systems, and implementation assessment would be appropriate methodologies here.

The questions and methods addressed under summative evaluation include the following:

- *What type of evaluation is feasible?* Evaluability assessment can be used here, as well as standard approaches for selecting an appropriate evaluation design.
- *What was the effectiveness of the program or technology?* One would choose from observational and correlational methods for demonstrating whether desired effects occurred, and quasi-experimental and experimental designs for determining whether observed effects can reasonably be attributed to the intervention and not to other sources.
- *What is the net impact of the program?* Econometric methods for assessing cost effectiveness and cost/benefits would apply here, along with qualitative methods that enable you to summarize the full range of intended and unintended impacts.

Clearly, this introduction is not meant to be exhaustive, but the evaluation perspective presented here is very consistent with the current research synthesis and evidence-based practice paradigm, as you will see subsequently.

16-8 The Rise of Research Synthesis, Meta-Analysis, and Evidence-Based Practice in Medicine, Social Science, and Education

In one of his early films, *Sleeper*, Woody Allen played a person who wakes up in the future after years of slumber. Among his first experiences upon waking was hearing a radio report proclaiming that a new study had shown that red meat was extremely good for you, the opposite, the announcer said, of what had long believed to be true. We haven't quite heard that report yet, but we have all had the experience of hearing or reading about some surprising new study that reports something different than what had been "common knowledge" until that point.

Woody's film and our everyday experience of news reports suggests that progress in research is a sort of all-or-nothing process in which truth is revealed by the latest *p* value but also suggesting that we shouldn't have much faith in the durability of the individual products of science. In the preface of their book, *Summing Up*, a classic in the field of research synthesis, Light and Pillemer (1984) cited a typically incisive comment from Mark Twain that captures the problem: "The thirteenth stroke of a clock is not only false itself, but casts grave doubts on the credibility of the preceding twelve" (p. viii). Thus, to counter the sense of doubt we need a valid method of taking all of the evidence into account; that is, to synthesize prior research.

As you begin to research the prior knowledge base in your topic area, there is a reasonable chance that you will come across "contradictory findings" that others have used to motivate their studies, and it is also reasonably likely that at the end of the study, further studies are advised. Once again, from this perspective it is reasonable to see research as a pretty disappointing enterprise, short on definitive answers and long on contradictions and controversies.

This situation may reflect the inevitable pace of an evolutionary process based on deliberate but flawed methods. But researchers going all the way back to Karl Pearson knew that some kind of summative accounting of data must be possible and that such an analysis would help overcome the inevitable limits of the individual study in terms of sampling, measures, and conclusions. It took until the latter part of the 1900s for research synthesis, meta-analysis, and evidence-based practice to come of age, in part because of controversies over some of the most basic practices in research.

16-9 Problems with Null Hypothesis Significance Testing

Kline (2004) recently wrote a text to help us move beyond the controversies about null hypothesis significance testing. The controversy essentially has revolved around the over-reliance on null hypothesis testing, particularly the misuse of probability values associated with statistical tests. As Kline and others (for example, Wilkinson and the Task Force on Statistical Inference, 1999) have noted, a *p* value indicates only the probability of an observed result, *assuming that the null hypothesis is true*. It does not "prove" the null or alternative hypothesis, it does not provide a measure of the likelihood of the result being replicated, and perhaps most critically, it does not tell us how big the result is or how important it is in practical or clinical terms. And as you know from our prior discussion of statistical power, a significance test result has much to do with sample size as well as the size of the effect. A large enough sample will always produce a statistically significant result. To more accurately report on the results of a study, we should report effect sizes with

confidence intervals in order to gauge the precision of the estimates and should provide a contextual interpretation that includes attention to practical and/or clinical significance as much as possible.

16-10 Effect Sizes and Confidence Intervals

The fundamental idea of meta-analysis and research synthesis is the **effect size** (Lipsey & Wilson, 2001). An effect size is a signal-to-noise ratio that expresses the size of a relationship or a difference in a standardized way. By translating findings from different studies into a common metric, summary effect sizes (weighted by sample sizes and taking into account methodological strengths and weaknesses) can be calculated. These effects might be represented in a statistic such as Cohen's *d*, which reflects the standardized mean difference between groups across multiple studies. Or an effect size might be constructed using the correlation or variance accounted for. No matter the index of effect uses, it is very important to calculate and report a confidence interval to give a sense of the precision of the effect size. The confidence interval allows the reader to know the likely upper and lower bounds of the estimated effect.

effect size
An effect size is a signal-to-noise ratio that expresses the size of a relationship or a difference in a standardized way. Effect sizes are important in planning studies so that we can estimate sample sizes accurately, in comparing results across studies, and in synthesizing research results as in meta-analysis.

16-11 Statistical, Practical, and Clinical Significance

Commentators from medicine, social science, and education have all recognized the importance of providing effect sizes and confidence intervals along with a contextual interpretation, and different fields have adapted to the circumstances of their problems and populations. For example, in medicine, an effect size might be expressed in an index known as the *Number Needed to Treat (NNT)*, that is, the number of patients a doctor has to treat in order to expect a "benefit" to be observed. The NNT is considered an index of clinical significance because it impacts clinical decision making. The benefit is often put into practically significant terms so that both patient and doctor can discuss such practical issues as the likelihood of being able to resume a vocation after treatment. Kline's book (2004) provides a detailed and very readable guide to computation of effect sizes and confidence intervals for most of the designs introduced in this text. But the determination of clinical and practical significance is very much in its infancy in most fields. As in our previous discussion of power analysis, researchers may use pilot data, prior published studies, and expert judgment to estimate clinically or practically significant effects.

16-12 Research Synthesis

Research synthesis, the process of producing integrative summaries of a body of research, has become a primary activity in science and applied social science. This is not only because of the oft-cited "explosive growth" in virtually every field but also because of a very practical issue: Before we plan a new study we should have a reasonable idea about where things stand. In the past, this was accomplished through the narrative literature review, a process now regarded as less systematic and more subjective than is possible or desirable. In addition, we need a systematic way to evaluate both the positives and the negatives in the history of a research problem.

Most of the published research syntheses are based on quantitative data and are specifically called *meta-analysis*. But there are qualitative synthesis methods as well, although there is less standardization in procedures, similar to the state of

research synthesis
A structured empirical process of producing integrative summaries of a body of research.

things in the primary qualitative research. First we will consider meta-analysis (quantitative synthesis), and then move on to qualitative and mixed methods.

16-12a Meta-analysis

meta-analysis
A structured empirical analysis that integrates and summarizes results of multiple research studies or evaluations.

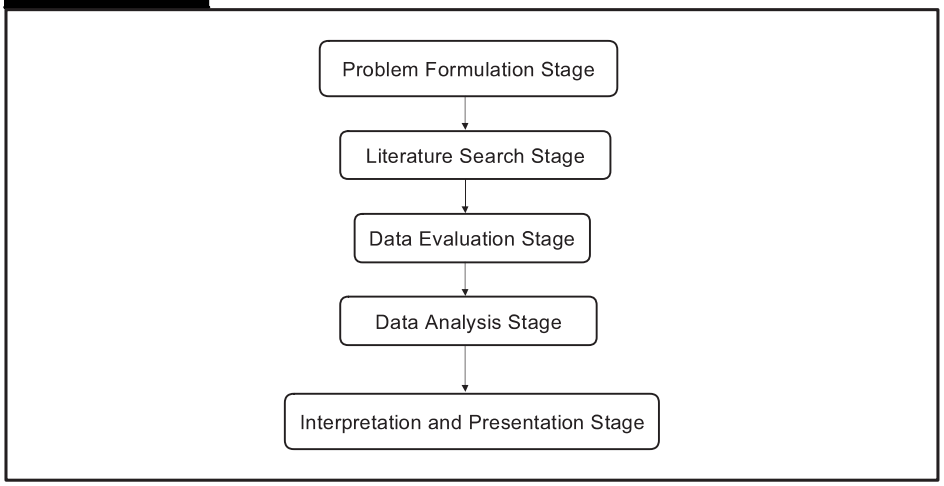
Cooper (1998) described a general stage model for **meta-analysis** or quantitative research synthesis. A similar model was later proposed by Lipsey and Wilson (2001) who added helpful details on quantitative procedures, including appendices with Effect Size formulas and Excel routines to conduct the analyses.

The basic stages of the model are shown in Figure 16–1. The most important and perhaps obvious characteristic of the model is that research synthesis turns literature review into a replicable scientific study. That is, instead of the subjective narrative review with variability in selection of studies, evaluation of evidence, and summary conclusions, we have a set of well-defined steps that clearly show how the synthesis of study findings was accomplished. That is not to say that the model makes research synthesis easy. In each stage, there are formidable challenges as an extensive methodological literature on the topic shows (Cooper & Hedges, 1994).

The first step of a research synthesis is to define the problem to be studied. As in a primary study, this may involve a statement of a hypothesis regarding a particular effect to be estimated (for example, Cooper has extensively examined the effect of homework on academic achievement) or may be more exploratory (for example, what is the best treatment for gambling addiction?). In either case, as in a primary study, the constructs of interest must be identified and defined. The synthesist will often encounter variations in definitions as well as the measures that have been used in prior studies. Decisions on which constructs and measures to include versus exclude must be documented, as well as the justification for such decisions.

In the literature search phase, the research synthesist encounters more decisions, including what studies to include in the review and how they will be accessed. Meta-analysts need to carefully specify inclusion criteria and document their efforts to explore the multiple channels available in study retrieval. Study inclusion criteria should be explained to the reader so that the tradeoffs between broad inclusiveness and a more restricted but higher quality set of studies are clear to the reader. The search procedures should be exhaustive and show efforts to retrieve studies via database searches as well as government documents, citation indexes, relevant journal, conference proceedings and consultation with well-known experts in the field (Lipsey & Wilson, 2001).

FIGURE 16–1 **Cooper's stages of research synthesis (1998)**



In addition, the development of a coding scheme is a major task. The coding scheme will enable systematic review of each study and typically includes study identification information, characteristics of the design such as type of design, sample, and measures, and detailed recording of all of the statistical outcomes reported. One of the key issues to be aware of in the search as well as the data evaluation stage is the "file drawer problem." It is well known that researchers are more likely to submit studies for publication if the results appear to be "significant," and it is also known that journal editors favor studies with "significant" results (Begg, 1994), again making the thoroughness of the search procedures a critical feature of the quality of the meta-analysis.

The data evaluation stage involves critical review of the yield of the literature search. In this phase, criteria may be applied that limit which kinds of designs will be included in the subsequent analysis. Cooper (1998) provided a set of possible guidelines for the critique, including a "threats to validity" perspective that is quite consistent with the emphasis of this text.

The next stage, data analysis, occurs when the analyst has extracted the statistical results from the studies and mathematically combines them into the overall effect size. This process involves assessing the variability of the observed effect sizes. If there is large variability (and there are statistics, most notably the Q statistic, to tell you), then consideration of what might account for that variability must be given. Quite often, such consideration includes study of potential moderators of effect sizes. Moderators are variables that interact with other study variables to influence outcomes. These may either have to do with the substantive area being studied or the designs used in the sample of studies. The data analysis stage has been facilitated by the emergence of computer programs. These include complete systems, such as *Comprehensive Meta-Analysis* (http://www.meta-analysis.com/), as well as guides such as Lipsey and Wilson's (2001), that provide examples of routines that can be run in common software such as Excel and SPSS.

The reporting stage should be handled as with any other scientific report. That is, sufficient information on all procedures and analyses should be included so that others may critically review or replicate the study. Graphic plots, such as the forest plot showing the effect size and confidence interval for all studies, are extremely helpful in communicating the results of a meta-analysis. Cooper also advised reporting on all possible threats to validity of study conclusions and his book provides numerous practical suggestions for assessing and reducing threats to validity of a research synthesis.

16-13 Qualitative Meta-Synthesis

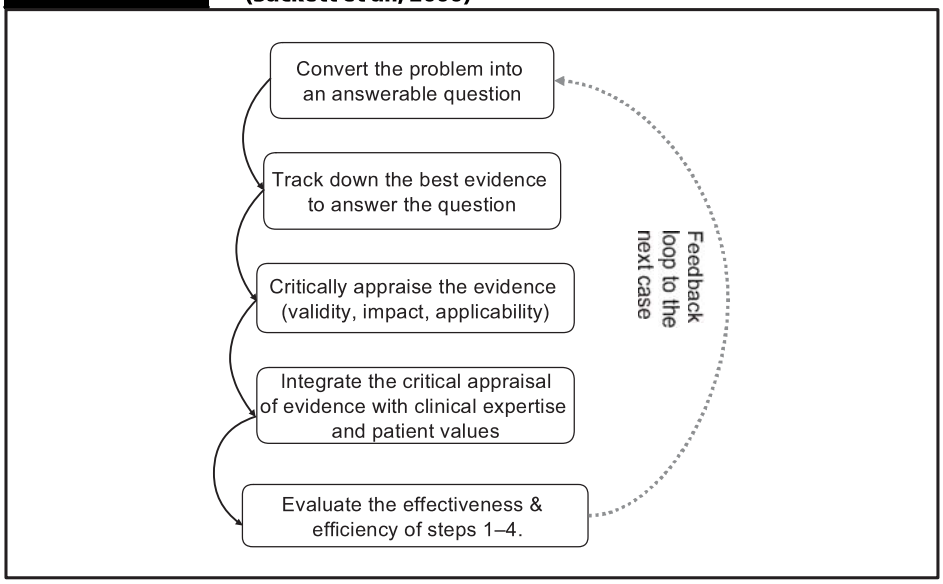
Most of what has been written about research synthesis has focused on quantitative studies. But there is increasing interest in developing methods for synthesizing qualitative research as well. For example, Jensen and Rodgers (2001) described case study research as a goldmine waiting to be discovered. Similarly, Stake (2006) has recently produced a text that shows a method of cross-case analysis. These and other qualitative synthesis methods are in an early stage of development. Perhaps they have most in common with content analysis, reviewed in Chapter 13. In addition, it is fair to say that mixed methods research is at its core a synthetic approach because in most cases the goal of the study is to integrate results across more than one method within the single study.

16-14 Evidence-Based Practice

Medicine, perhaps motivated by the tremendous stakes for individuals and society as a whole, has been central to the development of **evidence-based practice** based on research synthesis. Sackett and colleagues (2000) described the history of

evidence-based practice
Evidence-based practice means the use of the best available programs or treatments based on careful evaluation using critically reviewed research.

FIGURE 16–2 **Sackett and colleagues' steps in evidence-based medicine (Sackett et al., 2000)**



evidence-based medicine (EBM) in their concise guide. Beginning with the leadership of Professor Gordon Guyatt at McMaster University in Ontario, Canada, EBM has grown into a worldwide phenomenon, changing the culture of medical science, practice, and training profoundly. Sackett and colleagues defined EBM as "the integration of the best research with clinical expertise and patient values" (p. 1). Note that this is a three-part definition, giving weight to clinical judgment and patient preferences.

Sackett and colleagues developed the five-step model of EBM practice shown in Figure 16–2. The process was specifically designed for the clinical practice of medicine, but it can be generalized as a strategy for evidence-based practice. The first step is to put the problem into the form of an answerable question or questions. This means that the problem should be framed in terms that empirical inquiry may possibly help us with. Sackett and colleagues differentiated between background questions that address general aspects of a disorder such as causes and foreground questions that focus on the particular patient and possibilities for treatment, and they encouraged the evidence-based practitioner to include both kinds of questions.

The second step, tracking down the best evidence, has been transformed by computing technology and the efforts of professional groups to organize research synthesis so that it is widely accessible. Internet access to such databases means that in moments a doctor can find the most current research evidence, including synthesis, available. In addition to the main academic databases like *MedLine* and *PsychINFO*, there are evidence-based centers such at the *Cochrane Collection* in health care research (http://www.cochrane.org/) and the *Campbell Collaboration* for social science research (http://www.campbellcollaboration.org/), There are also specific evidence-based journals such as *Evidence-based Mental Health* that provide continuous updates on clinical problems.

The third step, critical appraisal of the evidence, is where the clinician's expertise as both a scientist and practitioner really kick in. At this step, Sackett and colleagues advise that the found evidence be appraised for three aspects: (1) validity, the apparent truth of the evidence based on the strength of the research; (2) impact, or the size of the effect reported in the research; and (3) applicability, or utility for the particular patient and setting.

Next, the clinical-scientist assesses the patient's values and preferences, discussing possible outcomes of treatments and side effects. In fields such as education, this may involve consultation with parents and other teachers and school personnel regarding a child's personality, abilities, interests, home situation, and so on. The relevant data on the individual case is then combined with the research evidence via a clinical judgment process in which parameters for a trial of a particular treatment, including how the outcome will be evaluated, are established.

Finally, the implementation and results are evaluated, with the goal of not only assessing the current status of the individual, but also feeding back into the cycle of evidence-based practice. In this sense, the evidence includes both that garnered from the research literature and from the lived experience of the practitioner with the patient, student, client, or other.

16-15 Conclusion: An Evaluation Culture

The ideas and methods introduced in this chapter suggest an important and implicit aspect of applied social research, the capacity for reflection and self-evaluation. Upon reflection, it is clear that we exist in a community of social researchers, a culture of its own. What are the values of the culture? Or, perhaps more important, what *should* its values be? Here I provide a vision of the types of values I would like to see become an integral part of 21st century thought in general, and of the applied research community in particular. There is no special order of importance to the way these ideas are presented; I'll leave that ordering to subsequent efforts. As you read this discussion, ask yourself how similar these values are to your own. What values do you think the applied research community should adopt? Here are my views.

First, an evaluation culture will embrace an *action-oriented* perspective that actively seeks solutions to problems, trying out tentative ones, weighing the results and consequences of actions, all within an endless evolutionary cycle of supposition-action-evidence-revision that characterizes good science and good management. This activist evaluation culture will encourage innovative approaches at all levels. However, well-intentioned activism by itself is not enough and may at times be risky, be dangerous, and lead to detrimental consequences. In an evaluation culture, you won't act for action's sake; you'll always attempt to assess the effects of your actions.

This evaluation culture will be an accessible, *teaching-oriented* one that emphasizes the unity of formal evaluation and everyday thought. Most evaluations will be simple, informal, efficient, practical, low-cost, and easily carried out and understood by nontechnicians. Evaluations won't just be delegated to one person or department; everyone will be encouraged to become involved in evaluating what they and their organizations do. Where technical expertise is needed experts will be encouraged to also educate others about the technical side of what they do, trying to find ways to explain their techniques and methods adequately for nontechnicians. Considerable resources will be devoted to teaching others about evaluation principles.

An evaluation culture will be *diverse, inclusive, participatory, responsive,* and *fundamentally nonhierarchical.* World problems cannot be solved by simple silver-bullet solutions. There is growing recognition in many arenas that the most fundamental problems are systemic, interconnected, and inextricably linked to social and economic issues and factors. Solutions will involve husbanding the resources, talents, and insights of a wide range of people. The formulation of problems and potential solutions needs to involve a broad range of constituencies. More than just research skills will be needed. Especially important will be skills in negotiation and consensus-building processes. Evaluators are familiar with arguments for greater diversity and inclusiveness; they've been talking about stakeholder, participative, multiple-constituency research for nearly two decades. No one that I know is seriously

debating anymore whether there should be a move to more inclusive participatory approaches. The real question seems to be how such work might best be accomplished, and despite all the rhetoric about the importance of participatory methods, there is a long way to go in learning how to accomplish them effectively.

An evaluation culture will be a *humble, self-critical* one. Researchers will openly acknowledge limitations and recognize that what is learned from a single evaluation study, however well designed, will almost always be equivocal and tentative. In this regard, I believe cowardice in research is too often undervalued. I find it wholly appropriate that evaluators resist being drawn into making decisions for others, although certainly evaluation results should help inform decision makers. A cowardly approach helps prevent the evaluator from being drawn into the political context, helping ensure the impartiality needed for objective assessment, and it protects the evaluator from taking responsibility for making decisions that should be left to those who have been duly authorized and who have to live with the consequences. Most program decisions, especially decisions about whether to continue a program or close it down, must include more input than an evaluation alone can ever provide. While evaluators can help elucidate what has happened in the past or might happen under certain circumstances, it is the responsibility of the organization and society as a whole to determine what ought to happen. The debate about the appropriate role of an evaluator in the decision-making process is an extremely intense one right now in evaluation circles, and my position advocating a cowardly reluctance of the evaluator to undertake a decision-making role may well be in the minority. This issue needs to be debated vigorously, especially for politically complex, international-evaluation contexts.

An evaluation culture will need to be an *interdisciplinary* one, doing more than just grafting one discipline onto another through constructing multidiscipline research teams. Such teams are needed, of course, but I mean to imply something deeper, more personally internalized—a need to move toward being nondisciplinary, consciously putting aside the blinders of peoples' respective specialties in an attempt to foster a more whole view of the phenomena being studied. As programs are being evaluated, it will be important to speculate about a broad range of implementation factors or potential consequences. It should be possible to anticipate some of the organizational and systems-related features of these programs, the economic factors that might enhance or reduce implementation, their social and psychological dimensions, and especially whether the ultimate utilizers can understand or know how to utilize and be willing to utilize the results of evaluation work. It should also be possible to anticipate a broad spectrum of potential consequences: system-related, production-related, economic, nutritional, social, and environmental.

This evaluation culture will also be an honest, *truth-seeking* one that stresses accountability and scientific credibility. In many quarters in contemporary society, it appears that people have given up on the ideas of truth and validity. An evaluation culture needs to hold to the goal of getting at the truth while at the same time honestly acknowledging the revisability of all scientific and research-based knowledge. It is important to be critical of those who have given up on the goal of getting it right about reality, especially those among the humanities and social sciences who argue that truth is entirely relative to the knower, objectivity an impossibility, and reality nothing more than a construction or illusion that cannot be examined publicly. For them, the goal of seeking the truth is inappropriate and unacceptable, and science a tool of oppression rather than a road to greater understanding. Philosophers have, of course, debated such issues for thousands of years and will undoubtedly do so for thousands more. In the evaluation culture it will be important to check in on their thinking from time to time, but until they settle these debates, it is necessary to hold steadfastly to the goal of getting at the truth—the goal of getting it right about reality.

This evaluation culture will be prospective and *forward looking*, anticipating where evaluation feedback will be needed rather than just reacting to situations as

they arise. Simple, low-cost evaluation and monitoring information systems will be constructed when new programs or technology are initiated; it will not do to wait until a program is complete or a technology is in the field before turning attention to its evaluation.

Finally, the evaluation culture I envision is one that will emphasize fair, open, *ethical, and democratic* processes. This will require moving away from private ownership of and exclusive access to data. The data from all evaluations needs to be accessible to all interested groups, allowing more extensive independent, secondary analyses and opportunities for replication or refutation of original results. Open commentary and debate regarding the results of specific evaluations should be encouraged. Especially when multiple parties have a stake in such results, it is important for reporting procedures to include formal opportunities for competitive review and response. An evaluation culture must continually strive for greater understanding of the ethical dilemmas posed by research. The desire for valid, scientific inference will at times cause conflicts with ethical principles. The situation is likely to be especially complex in international-evaluation contexts where evaluations may involve multiple cultures and countries that are at different stages of economic development and have different value systems and morals. It is important to be ready to deal with potential ethical and political issues posed by research methodologies in an open, direct, and democratic manner.

Do you agree with the values I'm describing here? What other characteristics might this evaluation culture have? You tell me. There are many more values and characteristics that ought to be considered. For now, the ones mentioned previously, and others in the literature, provide a starting point for the discussion. I hope you will add to the list, and I encourage each of you to criticize these tentative statements I've offered about the extraordinary potential of the evaluation culture that is in the process of evolving today.

Summary

And so we come to the end of our journey, at least for this volume. This chapter takes a view from the mountaintop back at the research path we follow in a particular study and at the aggregate of paths made up of the many studies that we accomplish collectively as our knowledge evolves. We looked at the general idea of synthesizing across multiple research studies, and the specific quantitative approaches to meta-analysis of results. We then looked at the connection of research to practice in the recent development of the evidence-based practice movement. We introduced the idea of evaluation and evaluation research both as an endeavor for conducting a practical research study that can be utilized and to generate multiple studies for later synthesis. Finally, we sketched out the value system that might underlie a culture of evaluation within which research methodology is a central component. We wish you the best on your own research journeys and hope we have provided you with the basics you will need to accomplish them successfully.

Login to the Online Edition of your text at www.atomicdog.com to find additional resources located in the Study Guide at the end of each chapter.