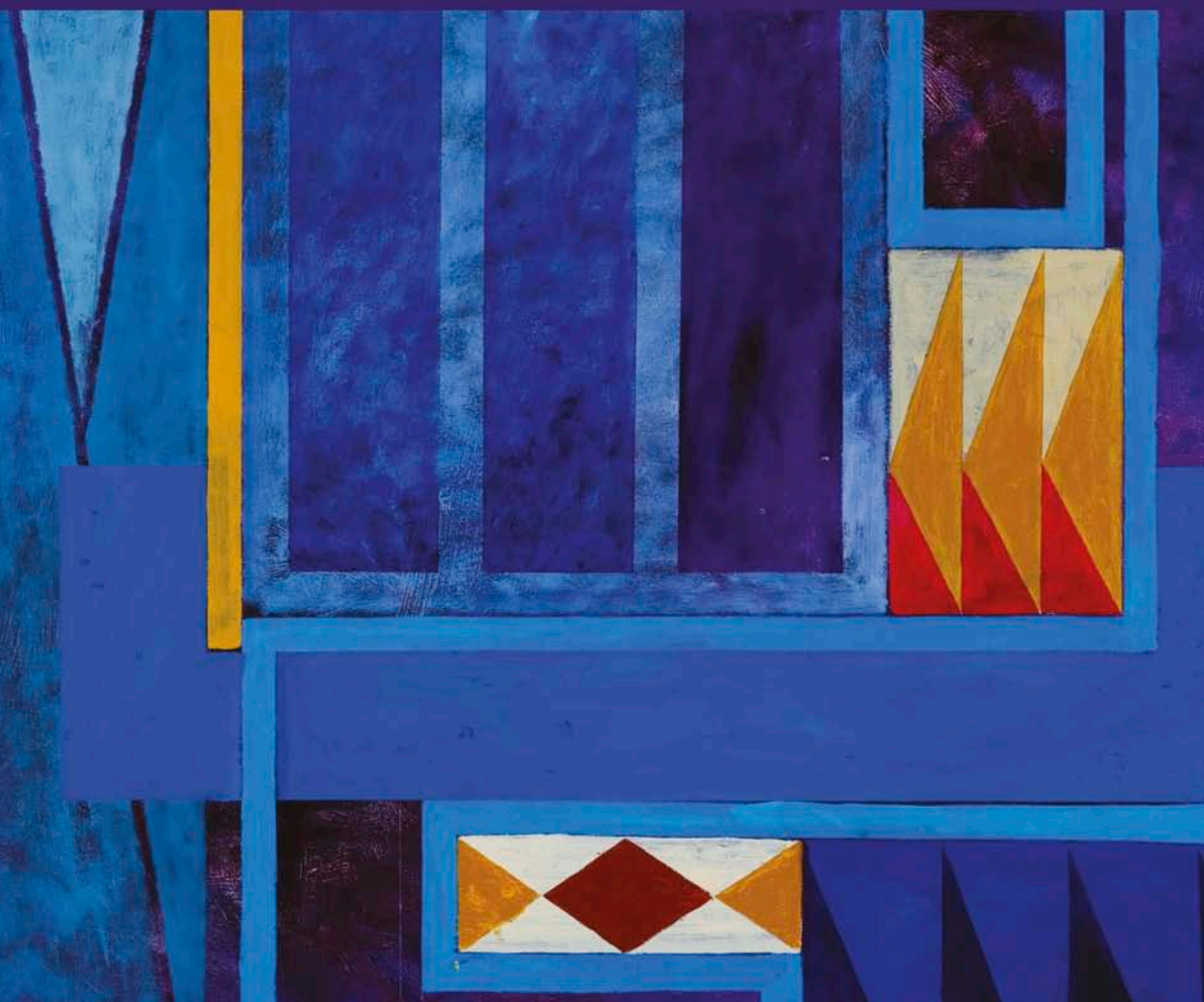
Program Evaluation

An Introduction to an Evidence-Based Approach

SIXTH EDITION

David Royse ▪ Bruce A. Thyer ▪ Deborah K. Padgett



Program Evaluation

An Introduction to an Evidence-Based Approach

Program Evaluation

An Introduction to an Evidence-Based Approach

SIXTH EDITION

David Royse
University of Kentucky

Bruce A. Thyer
Florida State University

Deborah K. Padgett
New York University

CENGAGE Learning®

Australia • Brazil • Mexico • Singapore • United Kingdom • United States

***Program Evaluation: An Introduction to an Evidence-Based Approach*, Sixth Edition**
David Royse, Bruce A. Thyer, and Deborah K. Padgett

Product Director: Jon-David Hague

Product Manager: Gordon Lee

Content Developer: Christopher Santos, J. L. Hahn Consulting Group

Product Assistant: Stephen Lagos

Media Developer: John Chell

Marketing Manager: Shanna Shelton

Manufacturing Planner: Judy Inouye

IP Analyst: Deanna Ettinger

IP Project Manager: Brittani Hall

Art and Cover Direction, Production Management, and Composition: Lumina Datamatics, Inc.

Content Digitization Project Manager: Elizabeth Cranston

Cover Image: © Ignacio Auzike/ Glow/Getty Images

Library of Congress Control Number: 2014957794

Student Edition:
ISBN: 978-1-305-10196-8

Cengage Learning
20 Channel Center Street
Boston, MA 02210
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at **www.cengage.com/global.**

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

To learn more about Cengage Learning Solutions, visit **www.cengage.com**.

Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com.**

Printed in the United States of America
Print Number: 01 Print Year: 2014

Contents

Preface

A textbook must provide, first and foremost, information to assist the reader in better understanding the topic. The information must be credible and easily accessed and digested. Readers of this edition will find numerous examples from the literature to illustrate key points and concepts. You will discover a new emphasis on evidence-based practice and we continually stress the importance of locating evidence on the effectiveness of programs and policies. We believe that program evaluators and graduate students learning how to evaluate must be acquainted with salient and contemporary evaluations of programs similar to those they are interested in. This can provide useful information on effective and ineffective programs and practices, as well as examples of the types of designs used to evaluate similar programs with the latter providing guidance useful for crafting one's own program evaluation.

Textbooks that have gone through multiple editions continue to improve as a result of reviewers' comments and readers' feedback, and this one is no exception. Looking back over the efforts associated with the sixth edition, we have built upon the solid foundation of previous editions, but added new material as well.

Sometimes it is hard to know how much to "redo" and how much to leave alone. Salient and classic studies are important, but there are always new ideas and recent studies to highlight. Some content must always change to keep a text current. However, educators choose books they have used in the past because they know and like that content. If a book changes too much, then it may not always have the same comfortable fit as when it was previously chosen. It is somewhat of a challenge, then, to know what to keep and what to toss. And that is why, dear reader, you are invited to let us know what works and what doesn't. If there is some aspect of program evaluation that we don't explain very well, some chapter that doesn't work for you, then send an email and tell us what you think.

For those of you who are just discovering this book, *Program Evaluation* is designed to be a primary textbook for graduate students in social work, sociology, psychology, public administration, counseling, education, nursing, and other related disciplines where it is important to know if interventions, be they individual, family, group, or community focused, are having the desired impact. We hope that we have written clearly enough that practitioners, administrators, and

other persons who have just acquired the responsibility for overseeing or evaluating specific programs or policies may find it to be a valuable resource in planning evaluation or need assessment projects.

Our focus has been on communicating the essentials—that is, the basic tools and knowledge necessary to conceptualize a program evaluation and carry out the tasks associated with examining and appraising program performance. This book is designed to help students and practitioners understand *and* contribute to evidence-based practice. It is vitally important that we professionals and professionals-in-training continually examine the effectiveness of our programs and their impact upon our clients. This is the way improvement comes about—with benefits for both clients and society. We hope in some small way that our book is able to prepare and assist you for this important responsibility.

David Royse (**droyse@uky.edu**)
Bruce Thyer (**bthyer@fsu.edu**)
Deborah Padgett (**deborah.padgett@nyu.edu**)

Program Evaluation

An Introduction to an Evidence-Based Approach

CHAPTER 1 Introduction

Why is it important that practitioners be familiar with, and in some circumstances, be involved in, conducting program evaluation studies?

It is through program evaluations that services can be credibly shown to be helpful, ineffective, or harmful. It is the hallmark of a genuine profession that concern be given to the evidence-base of its services. All of the human service fields support the important of program evaluation and commends their members to take into account current credible outcome studies when arriving at practice and policy decisions.

What skills will I need?

All human service professionals need to be able to read, understand and critically evaluate program evaluation studies, and decide whether to apply such findings to their own settings. Some practitioners will find it necessary to help in the design, conduct and interpretation of program evaluations, as suggested by contemporary ethical codes.

THE IMPORTANCE OF PROGRAM EVALUATION

Welcome to the field of program evaluation, that aspect of professional training aimed at helping you to integrate research and practice skills, using the former to enhance the latter. We recognize that relatively few practitioners in the human services of social work, psychology, counseling, education, psychiatry, public health, public administration, and nursing will have careers primarily devoted to scientific research. However, every professional in these fields will be concerned with providing services to clients at some level (individuals, families, couples, small groups, organizations, or communities), and every human services practitioner will be intensely interested in learning whether the services provided really do help the clients they are

intended to serve. Thus, even if research may not be your major professional activity, learning about genuinely effective programs, and evaluating services and programs using scientifically credible research tools is a professional skill that you will find valuable. And apart from possibly being involved in the design and conduct of program evaluations yourself, you will read and critically analyze evaluation studies conducted by others and published in various disciplinary and interdisciplinary journals, books, and websites. Your ability to judge the value of published and unpublished program evaluations is a valuable skill to acquire. Otherwise, how will you be able to decide what types of services are worth providing to clients, at either individual or programmatic levels?

In the human services fields, scientific research can be broadly classified into having three main (often interrelated) purposes: to objectively describe things; to empirically evaluate the effectiveness of services; or to validly explain things. Descriptive research can be undertaken to better understand the characteristics or needs of clients of a particular agency. Evaluative research helps determine whether these needs are being met or clients' goals attained, while explanatory studies aim at uncovering the causes of psychosocial problems or the processes by which interventions work. This latter endeavor may contribute to what is known respectively as etiological or interventive theory.

Program evaluation can be seen as a subset of those activities labeled research, which itself has been simply defined as "systematic procedures used in seeking facts or principles" (Barker, 2014, p. 365). A more expansive definition of research is displayed in Table 1.1, provided by the American Educational Research Association. It is worth emphasizing the points found in this definition. Research must be rigorous, systematic and seeks objective information. A genuinely scientific study cannot be conducted casually, without careful planning, and must include a dedication to find out *what actually is*, as opposed to attempting to prove one's preexisting expectations. It must be logical, and a very wide range of methods can be employed, methods covered in the balance of this book. The term *instruments* in evaluation research, usually refers to pencil and paper scales completed by clients or other stakeholders in a program (e.g., caregivers, supervisors, teachers, etc.). Sometimes program evaluation involves the direct measurement of observable behavior, in which case sound studies include appropriate methods to assess the reliability and validity of these observational methods. Less frequently, program evaluations may include the use of some sort of physical instrumentation, such as the use of biofeedback devices, blood pressure or heart rate monitors. And occasionally program evaluations assess clients' physiological variables such as blood sugar, cholesterol levels, or measures of "stress."

Many program evaluations involve the analysis of outcomes using various forms of descriptive and inferential statistics, and such methods must meet the underlying assumptions of these tests and be carried out correctly. When given a choice of different legitimate approaches, it is wisest to use the more conservative methods of analysis, as this reduces the likelihood of finding reliable but pragmatically weak effects or results. Research reports must be clearly written, so clear that in theory, if not in fact, an independent well-trained investigator could replicate the essential features of a given program evaluation. By doing so, subsequently finding similar positive results enhances the confidence that the original study's conclusions were legitimate.

Table 1.1 Definition of Scientifically Based Research*

I. The term *principles of scientific research* means the use of rigorous, systematic and objective methodologies to obtain reliable and valid knowledge. Specifically such research requires:

   A. development of a logical, evidence-based chain of reasoning;
   B. methods appropriate to the questions posed;
   C. observational or experimental designs and instruments that provide reliable and generalizable findings;
   D. data and analysis adequate to support findings;
   E. explication of procedures and results clearly and in detail, including specification of the population to which the findings can be generalized;
   F. adherence to professional norms of peer-review;
   G. dissemination of findings to contribute to scientific knowledge; and
   H. access to data for reanalysis, replication, and the opportunity to build on findings.

II. The examination of causal questions requires experimental designs using random assessment or quasi-experimental or other designs that substantially reduce plausible competing explanations for the obtained results. These include, but are not limited to, longitudinal designs, case control methods, statistical matching, or time series analysis. This standard applies especially to studies evaluating the impacts of policies and programs on educational outcomes.

III. The term *scientifically based research* includes basic research, applied research, and evaluation research in which the rationale, design, and interpretations are developed in accordance with the scientific principles laid out earlier …

*Obtained from the American Educational Research Foundation website, found at: **http://www.aera.net/AboutAERA/KeyPrograms/EducationResearchandResearchPolicy/AERAOffersDefinitionofScientificallyBasedRes/tabid/10877/Default.aspx**

A defining characteristic of scientific research is that it is published in credible outlets, which make use of peer-review. This involves submitting the paper to a journal or other outlet, and having it scrutinized by two or more competent persons, which provide critical analysis and a recommendation to revise, reject, or accept the paper. This is usually done blindly, with the reviewers not knowing the identities or affiliations of the authors, and the authors not knowing the identities of the reviewers. This blind review is one way to try and reduce bias, so that reviewers are not influenced by an author's existing reputation, or the prestige of the authors' affiliations (e.g., Harvard versus Midwest A & M). Also the reviewers can provide more honest feedback if they are assured the authors will not know who they are (and perhaps "retaliate" in the future). Only after such rigorous vetting do scientific papers make it into the professional literature, and it is this layer of peer-review, which lends additional credibility to these studies. Professional journals (online or print) remain a major outlet for the dissemination of scientific research, although increasingly websites are used for this purpose.

In many instances, program evaluations look only at outcomes, which is a very legitimate undertaking. Other approaches attempt to determine client's perspectives,

such as their satisfaction with services, or how respectfully they were treated by agency staff. More ambitiously, some evaluation studies attempt to determine whether or not a given intervention actually *caused* any observed outcomes or differences between groups. Answering causal questions can be an exceedingly difficult undertaking and usually requires relatively sophisticated research designs. This book will cover the ranges of methods and research questions of interest to program evaluators.

Evaluation research refers to "systematic investigation to determine the success of a specific program" (Barker, 2014, p. 148). **Program evaluations** are a practical endeavor, not an academic exercise, and are not primarily an attempt to build theory or necessarily to develop social science knowledge (although it is wonderful when that happens). Tripodi (1987) has noted that "the mission of program evaluation in social work is to provide information that can be used to improve social programs" (p. 366). Similar definitions can be provided across the human service professions. For example, in *The Dictionary of Psychology*, Corsini (2002, p. 766) defines program evaluation as "… a process whose purpose it is to: (a) contribute to decisions on installing, certifying, and modifying programs; (b) obtain evidence to support or oppose a program; and (c) contribute to basic knowledge." Program evaluation is designed to make a difference in the lives of stakeholders, which can include clients, service providers, supervisor, administrators and policy-makers at all levels. One can evaluate the effects of a program at the level of an individual, or among many people who received a given program, at similarly constituted programs applied in various settings, and at the level of statewide, nation, or even international policies (e.g., human trafficking).

Curiously, although for many decades the singularly crucial importance of human services professionals designing and conducting program evaluations has been widely recognized, program evaluation research remains rarely undertaken, relative to the amount of literature published. Consider this fact in light of the quotations contained in Box 1.1.

Thus we have the very odd situation that although many authorities consider program evaluation to be perhaps the most valuable type of research contribution one can make, such studies seem to be very rarely undertaken and published. Instead, most empirical studies have a focus on descriptive and explanatory research work (Holosko, 2010; LeCroy & Williams, 2013; Rosen, Proctor & Staudt, 1999), which, at best, may have potential *implications* for practice in the human services, but not the potential for direct *applications* that a well-crafted outcome study would possess.

Why is it necessary to evaluate both new and established services? Because there are always alternative, and sometimes better, ways to solve problems. Abusive alcohol use is a serious problem on many college campuses and universities spend a good deal of money attempting to prevent and reduce alcohol abuse and its associated problems. A treatment known as Motivational Interviewing (MI) has been widely used as a treatment for alcohol abusers, and as a preventive intervention on college campuses. Recently Wilke, Mennicke, Howell, and Magnuson (2014) evaluated a MI prevention program provided to members of sororities and fraternities at a large university. A relatively strong evaluation design was used, which randomly assigned four Greek houses to receive the MI program *plus* the standard

BOX 1.1 Opinions on the Importance of Evaluation Research

I appeal to you.... Measure, evaluate, estimate, appraise your results, in some form, in any terms that rest upon something beyond faith, assertion, and "illustrative case." State your objectives and how far you have reached them.... Out of such evaluations will come, I believe, better service to the client. (Cabot, 1931)

The third type of research, evaluative studies of welfare programs and the activities of practitioners, are the most important of all. (Angell, 1954, p. 169)

[S]ocial work is not a science whose aim is to derive knowledge; it is a technology whose aim is to apply knowledge for the purpose of control. Therefore, on the research continuum social work research falls nearer to the applied end, because of its purpose of practical knowledge. (Greenwood, 1957, p. 315)

Evaluation and client feedback are not only necessary for effective service delivery, but are an ethical requirement of the profession. Systematic methods must be developed to assess whether social workers are helping, harming, or doing nothing for the people they serve. (Rosenberg & Brody, 1974, p. 349)

Social work has no more important use of research methods than assessment of the consequences of practice and policy choices.... [S]mall scale, agency-based studies are worthwhile if they succeed in placing interest in effectiveness at the center of agency practice and when they create a critical alliance between practitioners and researchers. (Mullen, 1995, pp. 282–283)

Studies are needed on the effectiveness of psychosocial intervention, including interventions previously tested under ideal controlled conditions, in real-world health care systems. (Ell, 1996, p. 589)

Research on actual service interventions is the critical element in connecting research to the knowledge base used by professional practitioners.... [T]he issue now is one of developing investigations of social work intervention initiatives, studies that go beyond descriptions and explanatory research. (Austin, 1998, pp. 17, 43)

We need to establish a research agenda for social work.... And intervention studies must be high in priority to such an agenda. (Rosen, Proctor, & Staudt, 1999, p. 9).

existing campus-wide alcohol abuse prevention efforts, and four to receive *just* the standard university services. Measures were taken of student drinking and of alcohol-related consequences, covering the 30 days before the program was implemented, and again 30 days afterward. It was hypothesized that students receiving MI plus usual campus services would reduce their drinking, and experience fewer negative consequences from their drinking, much more than the students receiving usual services alone. Sadly, this is not what was found. There were no differences attributable to the MI intervention on drinking or its negative consequences. There were no significant differences between the two groups at the pretest, nor at the posttest, when it was hypothesized that the positive effects of MI would be evident. Now, if you were a university administrator charged with locating and implementing a new campus-wide alcohol abuse prevention and intervention program aimed at college students, would you want to know about this article? Would the results

have any bearing on your choice of a new program to adopt? Perhaps so, especially if the MI lacked any counter-balancing studies, which had found that it *was* effective at the prevention and treatment of alcohol abuse on campus.

Mental health concerns afflict a sizeable number of individuals. A minority of these people seek or receive professional treatment, in part due to the lack of available licensed clinicians available to provide empirically supported treatments. Do mental health clients really require treatment from doctoral-level psychotherapists, or could less intensively trained clinicians also provide effective care? That was the question investigated by Stanley et al. (2014). Over 200 older adults with Generalized Anxiety Disorder (GAD) were randomly assigned to three conditions: Treatment by a doctoral-level psychotherapists; treatment by bachelors-level trained therapists; or to usually available care. Measures of GAD symptoms were taken after assignment to treatment conditions but before treatment began, and at pretreatment all groups were essentially equivalently disturbed and similar on demographic and other measures. The bachelors-level therapists were especially trained in a treatment known to be helpful for people with GAD, something called cognitive behavioral therapy (CBT). The same CBT intervention was provided by the doctorally trained clinicians. After six months, the clients were reassessed. It was found that the doctoral and bachelor-level therapists obtained similarly positive results with their clients, and both treatments outcomes were better than among those receiving usual care. These results suggest that mental health agencies may be able to provide effective services delivered via lower-cost (e.g., bachelors level) clinicians compared to doctoral-level psychotherapists. If you were a manager of a community mental health agency wanting to maximize your budget, would the results of this study be of interest to you?

Policy-makers of all political persuasions wish to know the impact of social policies. One major policy in the United States is called Medicaid, a health insurance program intended for relatively poor people. States are being encouraged to expand their Medicaid programs and the federal government has been providing financial incentives to induce them to do so. One argument made for Medicaid expansion is that it will improve peoples' health, and have some financial advantages such as reducing highly expensive emergency room (ER) visits for routine care. Poor people lacking health insurance often make inappropriate use of ERs because they can receive treatment there without necessarily having to pay for it. Providing Medicaid coverage to the poor, it has been argued, will give them the insurance coverage needed to obtain routine care from a primary care provider, such as a family medical doctor or nurse practitioner. Does Medicaid actually reduce ER use? That was the question posed by Taubman, Allen, Wright, Baicker, and Finkelstein (2014) in their study of the Oregon Medicaid experiment.

In 2008, Oregon was provided additional federal money to support expanding their Medicaid program. Not enough money was provided to cover everyone on the Medicaid waiting list (about 90,000 people), so Oregon used an unbiased lottery system to randomly select who would get Medicaid (about 30,000), and who would remain on the waiting list (about 60,000). ER visits to Portland, Oregon hospitals were reviewed over 18 months after this expansion of Medicaid. It was

hypothesized (and expected by most stakeholders involved) that ER use would be much lower among the patients who received access to Medicaid. The reality was that ER usage was about 40 percent *higher* among Medicaid recipients, compared to people who did not get Medicaid. Their conclusion?

> We find that expanding Medicaid coverage increased emergency department use across a broad range of visit types, including visits that may be most readily treatable in other outpatient settings. These findings speak to one cost of expanding Medicaid … and may thus be a useful input for informed decision-making balancing the costs and benefits of expanding Medicaid. (Taubman et al., 2014, p. 4)

Now imagine you are a governor or legislator charged with trying to decide whether or not to expand Medicaid in your state. As this book is being written, the Taubman et al. (2014) study is being used in some conservative states to justify *not* expanding Medicaid, on the grounds of increased cost. This illustrates the complex tangle of making political and policy decisions. We can be quite sure that Taubman et al. did not undertake their study in order to provide ammunition for politicians to deny access to health insurance to the poor. They wanted to simply find out what the results were, and these results would have played a corresponding role among more progressive politicians to justify expanding Medicaid, had the results demonstrated lower ER use. No one is arguing that empirical research findings be the sole criterion for policy-making, but current program and policy evaluations, especially very well-done ones, should certainly be a part of the decision-making process.

There are many examples of evaluations with positive results that can be used to develop effective policies and programs by human service practitioners seeking guidance as to what types of interventions work well, or work better than others. Researchers in the United Kingdom and in Sweden undertook a systematic review of program evaluations of home repair programs and their impacts on residents' health. The focus of home repairs in this report involved enhancing insulation, improving home heating systems, double-glazing windows, and draft abatement (chalking). Over 39 such studies were found of varying quality, some of which were of very high quality indeed. Briefly stated, thermal enhancements to homes were associated with decreased respiratory illnesses and enhanced general health. Absenteeism from work and school was reduced, and thermally enhanced homes had more "usable space," which improved psychosocial functioning in families (e.g., they did not need to cluster in one room with a space heater) (see Thomson, Thomas, Sellstrom & Petticrew, 2013).

As another example, a group of Scandinavian researchers completed a systematic review of high-quality evaluation studies examining the effects of welfare-to-work programs. Their 122-page analysis was grounded in a review of 46 programs involving over 412,000 participants worldwide, although most such studies were conducted in North America. These welfare-to-work programs "… results in a slight increase in the likelihood that the unemployed person will find employment, that earnings will increase and that welfare payments will be reduced. They also reduce the likelihood that the unemployed person will remain on long-term welfare" (see **http://www.campbellcollaboration.org/lib/project/18/**).

If you are working in the field of helping the poor find jobs, you would be well-advised to obtain this systematic review and to become familiar with those approaches shown to be more effective than others. Another systematic review on the Campbell Collaboration website found that cognitive behavioral treatment is helpful to children suffering from post-traumatic stress disorder. Another review supported the positive effects of after-school parent involvement at improving children's academic performance. The websites of the Campbell Collaboration (**www.campbellcollaboration.org**) and the Cochrane Collaborations (**www.cochrane.org**) contain numerous systematic reviews, which represent some of the most credible research projects ever undertaken in a wide array of domains in the human services, including social welfare, criminal justice, education, health, and medicine. These are exceptional resources for program evaluators to be familiar with.

We believe that previously published and credible evaluation studies should be a major source of input into the design of an agency's programs. Our hope is that through reading this book, you will acquire greater skills in locating and critically evaluating such studies, and in designing and conducting empirical evaluations of your own practice and of the outcomes of the agency where you may be employed.

When we improve our programs and interventions by making them more effective and efficient, all those involved with or touched by the human service delivery system are affected. Consumers and their families may recover faster when we discover that one approach works better than another. Armed with information from the program evaluation, workers and managers can better treat and advocate for their clients—possibly making their own jobs more enjoyable and less frustrating. Ultimately, even taxpayers benefit. But let us back up a bit and discuss what constitutes a program.

WHAT IS A PROGRAM?

A program is an organized collection of activities designed to reach certain objectives. Let's consider the two main elements of this definition in depth. Organized activities—programs—are not a random set of actions but a series of planned actions designed to solve some problem. If there is no problem, then there is no need for programmatic intervention. So, programs are interventions or services that are expected to have some kind of an impact on the program participants. Barker defines a program as a "Relatively permanent organization and procedure designed to meet ongoing client needs (as opposed to a 'project,' which is more flexible and short term in scope" (2014, p. 338). This book includes evaluations aimed at a range of venues—long-term programs, time-limited projects, one-time group research designs, and of the effects of programs (including individual services) on particular clients. We cover both quantitative, qualitative, and mixed method approaches to program evaluation, as well as other important forms of evaluations such as needs assessments, client satisfaction studies, formative and process evaluations, and cost effectiveness and cost analysis investigations. Each is important and encompassed within the broad scope of evaluation work. To begin, we will take a look at what are some of the features of good programs.

CHARACTERISTICS OF "GOOD" PROGRAMS

Programs tend to have certain characteristics that help us identify them. First of all, programs tend to require **staffing**. A residential drug treatment facility, for instance, is going to need a lot of staff. It may even have a separate staff who run an aftercare or outpatient drug treatment program. The personnel of both programs may occasionally be asked to speak to high school students and groups in the community as part of the facility's drug education program. Staff may have their time allocated among several programs or dedicated to only one.

Second, programs usually have their own **budgets**. Because employing staff requires financial resources, programs sometimes can be identified by their budgets. However, some fine programs have minimal budgets because of heavy reliance on volunteers. **Stable funding** is important to the success of most programs. Morale and performance fall when employees do not get paid on a regular basis, or when they are asked to put aside normal duties and engage in last minute fund-raising or grant writing to get the program through several more months. Programs started with "soft money" (grants or nonrecurring funds) often experience high rates of staff turnover until the programs secure some continuity in funding.

Another characteristic of programs is that they have their own **identity**. In short, they are visible or recognizable by the public. Big Brothers Big Sisters (BBBS) is an example of an organization with a national reputation for a single program. Others you may recognize include Habitat for Humanity, Alcoholics Anonymous, or the National Alliance for the Mentally Ill. In some communities, a program may be recognized by the location where it has been housed for a number of years, or by its unique slogan, sign, letterhead, spokes person, or public service announcements.

When an organization has multiple programs, differences are sometimes found in philosophies, policies or procedures, and missions, and perhaps even in the way their corresponding staffs dress and how they account for their time. Such contrasts make it easy to differentiate one program from another. Within an agency, one outpatient counseling program may have the **service philosophy** that "no one is turned away," while another outpatient counseling program may have a different philosophy—providing service only for those who meet certain eligibility guidelines, such as having private insurance or being able to afford to pay. A service philosophy may also clearly communicate how the clientele is to be treated, for example, "We respect the dignity and worth of all those we serve in caring for their physical, spiritual, psychological, and social well-being" or "The customer is always right."

Unfortunately for program evaluators, programs can be vague and hard to distinguish and define. A former governor once made a public announcement that he was unveiling "a new program" to put state social workers in public schools. The program, he said, should help prevent dropouts and poor achievement among students who faced serious personal and family problems. However, the newspaper account said the program would require no additional staff or funds. In essence, some social services employees would be placed in schools that could supply them with office space and phone.

Did the governor's announcement create a program? Not in this instance. It never got off the ground. Why not? It had no name, no staff, no funds, no slogan,

no visibility. Most schools did not have surplus office space. Further, the governor made no suggestion of any new activities or ways of tackling the problems children and their families faced.

On the other hand, starting a bereavement support group in an elementary school, even if volunteers contribute the leadership and the group has no budget to speak of, could be considered a program if it has an ongoing presence and a presumed impact that could be measured. For evaluation purposes, speaking to an assembly of high school students once or twice a year about drugs and alcoholism might also be considered a program.

In the best of all possible worlds, every human services program would be solidly established on the basis of scientifically credible evidence that had been previously published in peer-reviewed professional journals. That is, before the practitioners jumped into a social problem and started "helping," someone did a serious search and appraisal of the relevant evaluation studies that tested the usefulness of various methods of potential helping. If a careful search of the literature and critical review of the existing outcome studies found that one or more models of intervention had credible evidence of effectiveness, and these approaches were "teachable" to the existing service providers, cost effective, and ethical, contemporary standards of ethical practice would suggest that the service program be focused around these empirically supported services as opposed to interventions lacking a sufficient foundation in empirical research.

Now, this poses a dilemma for practitioners and administrators, namely, "What if no empirically based services are known to exist for a particular problem?" In that case, one would be justified in primarily relying on the more traditional sources of practice knowledge, namely, *theory*, *practice wisdom*, *common sense*, *tradition*, and *authority*. But (and this is a big but), practitioners should only claim that no empirically based interventions exist after having made a thorough and up-to-date search of the relevant practice research literature. Fortunately, research-based interventions are now well established for a majority of the serious conditions described in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM; American Psychiatric Association, 2013), and increasingly for conditions that do not lend themselves to the DSM system—problems such as unemployment, domestic violence, child abuse and neglect, and troubled youth. We believe that every human services agency should keep abreast of these developments by subscribing to relevant journals (e.g., *Journal of Consulting and Clinical Psychology*, *Archives of General Psychiatry*, *Research on Social Work Practice*) and acquiring the latest edition of professional books that summarize the latest studies on research-supported practices.

Human services professional trying to locate credible information on the evidentiary status of various programs, policies, and services have a wide array of websites they can turn to. Table 1.2 lists a few of these.

Now of course because a program has been shown to be effective in the past, perhaps involving clients, therapists, and agency settings different from your own, there is no guarantee that such a program, if implemented in your agency or state, will yield similarly positive results. In philosophy this is known as the problem of **induction**. If you saw one new type of bird and it is white, and another, also white, and another, you may be tempted to conclude that all these birds are white.

Table 1.2 Selected Websites Containing Information on Research-Supported Psychotherapies and Psychosocial Programs and Policies

The Cochrane Collaboration (focuses on health care, broadly defined) **www.cochrane.org**
The Campbell Collaboration (focuses on social welfare, criminal justice, education, and social development) **www.campbellcollaboration.org**
Research-Supported Psychological Treatments (focuses on mental health disorders) **http://www.div12.org/PsychologicalTreatments/index.html**
SAMHSA's National Registry on Evidence-Based Practices and Programs **http://www.nrepp.samhsa.gov/**
California Evidence-Based Clearing House for Child Welfare **http://www.cebc4cw.org/**
National Institute for Health and Care Excellence (focused on Health and Mental Health) **http://www.nice.org.uk/GuidanceMenu/Conditions-and-diseases**
Agency for Healthcare Research and Quality—Evidence-Based Reports (focuses on Health) **http://www.ahrq.gov/research/findings/evidence-based-reports/overview/index.html**
Therapy Advisor—Consumer oriented website on research supported psychotherapies **http://therapyadvisor.com/defaultSub.aspx**
Coalition for Evidence-Based Policy—Social Programs That Work **http://evidencebasedprograms.org/**
What Works Clearinghouse—Focuses on Educational Interventions **http://evidencebasedprograms.org/**

That is commonsense. But the possibility remains that the next bird of the same type may be a different color. In program evaluation, if a program has been evaluated elsewhere and found to be helpful with clients with a particular problem, there is no guarantee that it will help similar clients with that problem in your agency. But if it has been repeatedly shown to be helpful with different types of clients, in differing agencies, with various service providers, and such studies have been conducted by independent investigators, and there is little to no counterfactual evidence to be found (no studies with negative results), then the rationale for choosing that intervention or program increases.

Practitioners and policy-makers *must* decide on what services to provide. Choosing from among programs with a stronger research-base enhances the likelihood that your clients will be benefitted. It does not guarantee success, but it is a surer guide than relying on alternative sources of information such as someone in authority telling you what to do, or relying solely on clinical intuition, theory, or prayer. Our perspective in this book is by all means make use of multiple sources of information in deciding upon what services to provide, but for heaven's sake make sure you consult

the recent, credible, scientific literature to find out what is already known about effective (and equally importantly, ineffective or harmful) services.

The question may legitimately arise, "How much evidence is enough evidence in order for a given intervention to be considered to have an adequate empirical foundation?" The *Task Force on Promotion and Dissemination of Psychological Procedures* (Chambless et al., 1996, p. 16) of Division 12 (Clinical Psychology) of the American Psychological Association has promulgated one set of minimal recommendations to address this question. In order for a psychosocial intervention to be considered for inclusion in their list of empirically supported treatments, it had to meet the following criteria:

The Treatment Must Be Supported By

1. At least two good between-group design experiments demonstrating efficacy in one or more of the following ways:
   a. Superior to pill or psychological placebo or to another treatment
   b. Equivalent to an already established treatment in experiments with adequate statistical power, or
2. A large series of single case designs ($N > 9$) demonstrating efficacy. These experiments must have:
   a. Used good experimental designs, and
   b. Compared the intervention to another treatment, as in 1a

Among the other criteria to be applied are that experiments must be conducted using treatment manuals (this enhances the ability to replicate interventions), the characteristics of the clients must be clearly specified, and effects must have been demonstrated by at least two different investigators or treatment teams. Although these standards may seem a bit stringent to students and practitioners unaccustomed to rigorous research, they are not unwarranted; and they serve as an initial starting place to begin classifying particular interventions as evidence-based or not. Human services professionals can employ this set of standards when selecting types of treatment programs to provide. Over time, it is likely that these standards will be augmented by additional criteria (e.g., "The treatment has been evaluated in real-life clinical settings") that will enhance their usefulness. Lists of psychosocial therapies which meet these standards can be found at the website on *Research Supported Psychological Treatments*—**http://www.div12.org/PsychologicalTreatments/index.html**.

Psychosocial treatments, programs, or policies, which enjoy a substantial degree of empirical support are variously referred to as empirically supported or research-supported treatments. What they should *not* be referred to as evidence-based practices. The following section will explain why.

EVIDENCE-BASED PRACTICE

The human services professions are slowly moving in the direction of evidence-based practice (EBP), which has been defined as "the integration of the best research evidence with our clinical expertise and our patient's unique values and circumstances" (Straus, Glazious, Richardson, & Haynes, 2011, p. 1). Although this statement may seem like common sense, the fact is that at present no clear

ethical or legal mandates require that human services professionals deliver research-supported interventions, where these practices are known to be established. This could change, however, in the near future. Clearly, it is the direction in which the human services professions are heading, and is being strongly encouraged by national governments in the United States, Canada, the United Kingdom, and other countries, and adopted by individual states in the United States. The definition of EBP quoted earlier comes from the fourth edition of the seminal book *Evidence-Based Medicine: How to Practice and Teach It.* Although EBP originated in medicine, its simple tenets have been widely adopted within many disciplines, including psychology, nursing, public health, counseling, and psychiatry. EBP outlines a basic five-step process practitioners can undertake to help them and their clients to decide what to do. These steps are:

1. Convert the need for information into an answerable question.
2. Track down the best evidence with which to answer that question.
3. Critically appraise that evidence for its validity, impact, and applicability to one's situation.
4. Integrate this critical appraisal with one's clinical expertise, and with the patient's unique values, circumstances, and preferences.
5. Evaluate one's effectiveness and efficiency in carrying out the preceding four steps.

The EBP model assumes that one follows appropriate professional ethical standards. While originally intended to guide physicians, the generic nature of these five steps were quickly adopted by other fields, and there are now large literatures dealing with the EBP practice model in various health and human services fields. Note that EBP is *not* a medical model. There is no assertion that client problems have their origins in the client's biology, no assertion that interventions must be medical, nor any contention that the clinician must be a physician (the three central principles of the medical model). Yes, EBP originated in medicine, but it is no more a medical model than the inferential statistics developed in agriculture in the early part of the 20th century implying that all the social and behavioral scientists who used them limits their use only to farmers!

Step 2, tracking down the available evidence can be conducted by an individual practitioner or program manager him- or herself, searching the literature, reading all of it, critiquing it, or one can locate useful summaries called ***systematic reviews*** **(SRs)** devoted to your answerable question, and take advantage of competent people already having done this preliminary legwork. If you can find a pertinent SR, this is a great savings of time and effort for you, and Table 1.2 lists some sources where SRs may be found, as well as citing and listing research-supported interventions. However a SR does not tell one what to do, what service to provide, or not to provide. A SR simply summarizes the existing evidence, and permits the reader to make use of it or not. For example, if a client needed a blood transfusion and her physician informed her of this, as a research-supported treatment for her condition, the client may refuse, perhaps on religious grounds. According to the model of EBP, the practitioner must respect the client's values, and not simply apply research-supported treatments without taking into account the client's views. Or perhaps a client was seriously depressed, and a recent SR indicated CBT was a well-supported psychotherapy but

if the client had a serious intellectual disability, the practitioner may opt *not* to provide CBT, given the judgment that the client was unable to undertake the self-reflection and homework required for this model. Instead, alternative therapies might be considered and discussed with client and/or responsible party, intervention such as behavior analysis, or antidepressant medication. In EBP there is *no* assertion that *research rules*! EBP does assert that research *must be considered*, along with client preferences and values, professional ethics, available resources and available skills. This is why, properly considered, there are no such things as evidence-based *practices* (nouns). There is the five step decision-making model of EBP, which is much more sophisticated than simply locating research-supported treatments and deciding to make use of them along with judicious consideration of the other equally crucial factors found in the EBP model (Thyer & Myers, 2011; Thyer & Pignotti, 2011). It is really quite important for program evaluators who wish to be informed about EBP to read the original sources of information about this model, such as Straus et al. (2011) and previous editions, and not on third or fourth-hand restatements, which often present quite distorted pictures of this model. See Thyer (2013) for an example of correcting misrepresentations of EBP, which have been presented in the social work literature.

The Role of Theory in Program Evaluation

In the best of all possible worlds, every program would also be based on a sound **theoretical model**. That is, before the practitioners jumped into a social problem and started helping, they would develop a model that would have examined the problem—how and why it originated and what would work best to remedy the situation. A psychosocial or other theory can be an organizing principle for each program, one that provides a consistency of effort by suggesting a standard approach derived from some well-articulated and comprehensive social or behavioral science theory. Such a theory may serve as a guide in conceptualizing the causes of problems and in proposed mechanisms of action for interventions.

Take the problem of hyperactive behavior (HB). If one approaches the understanding and treatment of children with HB from a biological orientation, one would focus on possible underlying disturbances in brain chemistry to explain HB and on the use of medications to regulate these presumed chemical irregularities. If one subscribed to a learning theory orientation, one might examine the possible role of an over-stimulating environment, or of peer/parental/teacher inadvertent reinforcement for HB, as tentative causes of HB, and of environmental manipulation and positive behavior management programs to reinforce on-task activities as an intervention. One trained in a biopsychosocial approach might employ both approaches. There are many theories on the causes of HB and theoretically based interventions, which could be considered.

Another example might be that of drinking too much alcohol (often called alcoholism). There are also a wide array of theories as to the etiology of abusive drinking, and of ways to help people drink less or to abstain from drinking. According to the theory of Alcoholics Anonymous (AA), alcoholism is caused by a biologically based allergic reaction to ethanol, in that the smallest sip triggers an irresistible craving to drink more, leading to a loss of control. This biological theory leads, naturally,

to an interventive theory, namely to abstain completely from exposure to alcohol. Just as a person with a peanut allergy is told to avoid all consumption of peanuts, the alcoholic is provided a program of education, interpersonal support, and spiritual development intended to reinforce their attempts to abstain completely. No one is ever said to "recover" from alcoholism or to be "cured," anymore than one is likely to cease having a peanut allergy. According to the AA theory—people are simply in various stages of "recovery." AA is a program of recovery clearly derived from both an etiological and interventive theory, in this case a biological one. However, there are certainly other theories of abusive drinking based on learning theory, genetic, psychodynamic, and familial influences, and so forth. Each theory has its own derivative programs (e.g., antabuse medication to produce nausea upon drinking, anti-craving drugs, behavior modification, the community-reinforcement approach to abstinence, psychotherapy, rational recovery, family therapy, meditation, etc.).

Consider a different problem. Suppose you are hired to run a treatment program for men who batter. Do these men fit a single profile? Is one interventive strategy all that is needed? Saunders (1992) argues that there are three theoretically distinct types of men who batter: those who were severely abused as children; emotionally volatile men with rigid sex-role attitudes, who fear losing their partners and are depressed, suicidal, and angry; and family-only aggressors who tend to have relatively liberal attitudes about sex roles, the lowest rate of abuse in childhood, the most marital satisfaction, and who are generally nonassertive. Is there a possibility that some interventions may work better with one type of abuser than with another?

The field of interpersonal violence, specifically spousal battering, is dominated by feminist approaches, which attempt to explain this problem in terms of the exploitative exertion of power encouraged by unfair patriarchal social institutions. This has led to the widespread adoption and use of the "Duluth Model" of intervention and prevention (see **http://www.theduluthmodel.org/about/index.html**). Alternative theories of spousal battering take a more gender-neutral approach, taking into account female-on-female and male-on-male battering in intimate relations (Ferreira & Buttell, in press). This is quite different from the Duluth Model. Thus, it can be seen how the theory one adopts for a given program can strongly influence how the problem is seen to be caused and how intervention may occur.

Theoretical models can be important to understanding how a program should work and where one should look for indications that a program is successful (see Box 1.2). But all too often what passes for theory in many social service agencies is a blend of past experience and tradition, and even myths or professional folklore.

BOX 1.2 Characteristics of "Good" Social Service Programs

- Staffing
- Budgets
- Stable funding
- Recognized identity
- Conceptual or theoretical foundation
- A service philosophy
- Systematic efforts at empirical evaluation of services
- Evidence-based research foundation

Evaluators would have no problem with that if the program was often successful in rehabilitating, helping, fixing, or curing clients. But when a program is not successful much of the time, the possibility exists that even though the program was implemented as designed, the underlying theory is flawed. Such a situation calls for new thinking and a new approach to the problem. Theories may be able to tell us how to accomplish our goals (Conrad & Miller, 1987), but not in every case.

Program evaluation studies should generally be considered as relatively poor tests of the validity of theories. Take, for example, a program with a theoretically derived intervention that is found to produce positive results with the clients—they appear to be really benefitting from the program's services. You might be tempted to conclude that this means that the program's underlying theory is thus proven to be valid. Doing so would be making a mistake. An intervention may work quite well for reasons completely unrelated to its underlying theoretical rationale (see Lilienfeld, Ritschel, Lynn, Cautin, & Latzman, 2014, for a comprehensive accounting of how this can happen). For example, the early (circa 1950s) behavioral psychotherapy called systematic desensitization (SD) was said to work because of an underlying theoretical process called reciprocal inhibition (RI). Careful tests in the late 1960s and 1970s found out pretty conclusively that while, yes, SD was moderately helpful in overcoming certain anxiety disorders it did *not* work because of any processes of RI. More recently, the psychotherapy called eye movement desensitization and reprocessing (EMDR, circa 1990s) was said to work because of underlying physiological changes in the brain created by asking the client to move their eyes back and forth in a certain pattern established by the therapist, while thinking of traumatic events. It turns out that these so-called saccadic eye movements are completely irrelevant to the modest improvements garnered by EMDR, which are more plausibly attributed to mild exposure to anxiety-evoking imagery and to placebo factors (see Lohr, Hooke, Gist, & Tolin, 2003). If someone improves after enrolling in a program based on behavior therapy, psychotherapy, or solution-focused treatment that does *not* prove that these respective theories are true. Generally speaking, the theoretical propositions that therapies and programs are based on require *evidence* of their validity independent of their successful outcomes.

Negative outcomes are also problematic in terms of making inferences about the validity of a program's underlying theory. Say you conduct a program evaluation of a theoretically based service and the outcome is negative, for example, the clients do not improve, or, heaven forbid, get worse. Doesn't this "prove" that the theory the program is based on is incorrect? Not so fast. Advocates of the theory may claim that the outcomes were negative because the treatment or service was not really implemented correctly, or that the therapists were insufficiently trained or supervised, the duration of treatment was too brief, the outcome measures were insufficiently sensitive to detect the delicate but far-reaching beneficial effects of therapy, that the follow-up period was not long enough, and so on. There can be a myriad of reasons (some genuine rationales and some spurious excuses) why a program may not work, and the invalidity of its underlying theory is but one. We do not encourage program evaluators to make inferences about the truth or falsity of theories based on the program's outcomes. In very rare cases, a stringently designed and executed outcome study with negative results *could* be seen as reflective of the underlying theory's invalidity, but it is much less credible to claim that a

positive outcome means that the theory is correct. Programs may work for many reasons apart from those hypothesized by the program's developers or conceptual grandparents.

To a certain extent, program designers and implementers are protected from many of the problems that accompany misguided and erroneous theories when they base programs not solely on theories alone but also on methods of practice that have been empirically supported by research and evaluation. The best programs, those on the "cutting edge," will be built on a firm foundation of the latest empirical research. These programs can provide a level of confidence and knowledge about what to expect in terms of success rates. These would be the programs we would choose for ourselves if we or our families needed intervention.

You should realize, however, that many (perhaps most) human services programs are *not* based on any explicit theory of human behavior or any etiological social or behavioral social science theory explaining how particular problems arise, or even any particular interventive theory. Such "atheoretical" programs may be based on common sense, authority, or tradition. For example, we know of one governor who arranged for the state to provide the mother of every newborn baby a compact disc (CD) of classical music. These CDs were duly distributed to thousands of new mothers, costing the state hundreds of thousands of dollars in taxpayer money. This authoritative initiative was instigated by the governor's reading about something called the "Mozart effect," wherein listening to a particular Mozart sonata supposedly resulted in college students earning higher test scores (not true). The (vain) hope was that mothers who were given CDs of classical music (which failed to contain a recording of the particular Mozart sonata in question!) would play these within hearing of their infants, who would grow up to be smarter. This human services program was not based on any etiological theory about intelligence, or even an interventive theory on how listening to classical music was supposed to improve intelligence. It was based on *authority*, the governor's mandate. Unfortunately, subsequent studies failed to demonstrate the Mozart effect in test taking, and none demonstrated any improvement in infant intelligence.

Numerous examples of such atheoretical programs can be found, from policies removing the driver's licenses of high school dropouts to policies within child protective service programs, homeless shelters, soup kitchens, and so forth. Ask the practitioners providing such services about their program's social or behavioral theoretical orientation, and often you will be met with puzzled looks. To be sure, these programs *can be* prospectively designed in accord with some theory, but in many instances they are not. And it can be misleading to *retroactively* attempt to fit a theoretical approach to explain a particular program, because the same program could no doubt be similarly explained by perhaps dozens of rival theories, leaving you no closer to a valid understanding of how the program may be exerting its effects, if any. The relationship between formal theory and research on practice outcomes is a complex one, and Thyer (1994; 2001) addresses it more completely than is possible here.

Not to be ignored are the possible *negative* effects of theory in human services programs, especially as they relate to programs and their evaluation (Thyer, 2012). Many social programs are likely based on false theory, theories based on invalid etiological and therapeutic propositions or pseudoscience (Pignotti & Thyer, 2009;

Thyer & Pignotti, 2010). So-called boot camps for juvenile delinquents are one example. Here the theory (loosely used) is that youth who get into trouble with the law have some sort of defect in their "character." Certain politicians' military experiences led them to believe that a Marine-style boot camp regimen could somehow correct the character of misbehaving young men (and women), and then, upon their release from the boot camp, be restored to their home communities having become somehow "stronger" in character and better prepared to resist the influence of evil peers. Well, the evidence is pretty clear that boot camps do *not* work well at preventing recidivism in youth (see systematic review on the topic available at **www.campbellcollaboration.org** and MacKenzie, Bierie, & Mitchell, 2007). Those states, which support boot camps for juveniles can be seen as providing a *dis*service to kids, squandering taxpayers' funds, and employing hundreds of staff engaged in delivering useless services, not to mention diverting youth from potentially more effective programs. The efforts expended by evaluators to appraise the outcomes of ineffective programs can also be seen as diverted from more useful activities. These are all harmful effects of grounding social services in invalid theories.

Programs vary greatly. Some are sophisticated and others simplistic—even composed of a single activity. In scale they range from a small cooperative babysitting program for young mothers to the federal food stamp program that touches millions of lives. It is not always easy to determine whether certain activities should be considered a program or part of the collection of activities that comprise a larger single program. Either way, program evaluation can be undertaken as long as a desired objective or outcome can be stated. Although in some agencies *programs* are referred to as *services*, in this book the terms will be used interchangeably.

PROGRAM EVALUATION DEFINED

Program evaluation is applied research used as part of the managerial process. Evaluations are conducted to aid those who must make administrative decisions about human services programs. Unlike theoretical research, where scientists engage in science for its own sake, program evaluation systematically examines human services programs for pragmatic reasons. Decision makers may need to know if a program accomplished its objectives, if it is worth funding again next year, or if a less expensive program can accomplish the same results. We like the following definition of program evaluation provided by Grinnell and Unrau (2008, p. 553):

> A form of appraisal, using valid and reliable research methods, that examines the processes or outcomes of an organization that exists to fulfill some social need.

Program evaluation is like basic research in that both follow a logical, orderly sequence of investigation. Both begin with a problem, a question, or a hypothesis. Normally, there is some review of what is known about the problem, including prior efforts and theoretical approaches (this is known as reviewing the literature). A research or evaluation design (a blueprint to guide the data collection efforts) is developed, and data are gathered and then analyzed. When thought of this way, both research and evaluation are similar to the task-centered or problem-solving process known to many human services professionals.

Research and evaluation differ with regard to the expected use or utility of the data. There may be no anticipated need or demand for "pure" research, whereas an assemblage of individuals may anxiously await the results of a program evaluation. Also, the goal of research is often to produce generalizable knowledge, while information from a program evaluation may be applicable to only a specific agency. However, both are approached with some degree of rigor. Think of program evaluation as a tool—a management tool that you can use to make (and to help others make) better decisions about social and human services programs. Program evaluation helps us to make the best use of our resources as we labor to improve the quality of life of our clients.

Program evaluation usually involves making comparisons. In fact, Schalock and Thornton (1988) have defined program evaluation as "structured comparison." Few programs can be evaluated without comparing them to something. Programs in one agency may be compared to similar programs in other agencies, to past or prior efforts, or against a stated objective; but without some form of comparison, there can be no evaluation. A major thrust of this book is to help you find or create (to conceptualize) bases of comparison for your own program evaluation efforts.

REASONS WHY PROGRAMS ARE EVALUATED

Quite often social and human services programs are evaluated because of a need to be accountable to a sponsoring or funding agency, or because competition for scarce funds requires that only one program (normally, the most effective or efficient program) can be funded. Program evaluation is needed whenever new interventions are being tried and it is not known whether they will be as successful as former methods, or when there is a perception that a program could be improved—that it could become more productive or better in some way. We evaluate on those occasions when it is important to have some objective assessment or feedback about the worth of our social and human services programs. The following scenarios illustrate some of the occasions when program evaluations are encountered.

Scenario 1: The Required Evaluation—Your agency is applying for funding from the United Way in your community to begin a new program designed to provide counseling to men who have been prosecuted for domestic violence. You have been asked to prepare the program proposal. As you read the instructions for preparing the proposal, you notice that besides describing the project, listing its objectives, pointing out its uniqueness, and stating the amount of funding that will be required, the proposal also requires a project evaluation. At the end of the project year, data must be presented to show that the project had a successful outcome and an impact on the problem of domestic violence.

Scenario 2: Competition for Scarce Funds—Your innovative program for men who batter has been operating for a year. You have been able to obtain some data that you hope will favorably influence the committee deciding the continuation of funding for your program. As you prepare your presentation, you discover that a second domestic violence project from another agency will also be making a request to be funded. You further learn that there is only enough money to fund one program.

Scenario 3: Evaluation of New Interventions—Many more clients desire the services of your outpatient counseling agency than you have staff to serve. At a planning session, one of the newer staff members suggests that the agency move from a one-on-one counseling model to a group services model. The benefits are clear—instead of limiting each practitioner to seven or eight scheduled clients a day, each therapist could conduct three or four group sessions a day and have contact with 25 to 30 clients. In spite of being able to serve more clients, the staff is not very supportive of this proposal, because they believe that individual counseling is much more effective than group counseling.

Scenario 4: Evaluation for Accountability—You work in a large residential agency serving young children. Unfortunately, a child care aide was recently discovered molesting one of the children. The public is in an uproar. Community leaders are calling for the agency director and all key staff to resign. You feel that the agency is a good one—better than other residential programs within the community. Because the agency director knows that you are enrolled in a program evaluation course at the nearby university, she calls you into her office and asks you to find some way of objectively documenting the strengths of the agency. "Can you show," she asks, "that the great majority of our young people have a favorable experience here, a good impression of the agency, and that they go on to do well in school and in life after they leave the agency?"

MOTIVATIONS FOR PROGRAM EVALUATION

Why do we evaluate human services programs? Programs are evaluated basically because administrative decisions have to be made, and it is important to know (or to show) that our programs are "good" programs. Individual policy or decision makers may have a hypothesis about a program (e.g., the free clinic's counseling program is highly effective). At other times, questions may be raised (e.g., is the free clinic's counseling program effective?). Hypotheses or questions provide the motivation for a program evaluation. It makes no real difference whether a question or a hypothesis serves as the catalyst for an evaluation. This can be seen in Box 1.3.

The list in Box 1.3 could easily be made much longer. An interest in exploring one question may lead to other areas. The evaluator may start off wanting to know whether clients were being helped, but in the process of designing a methodology the initial question or problem becomes modified somewhat. The evaluator may want to know not only whether clients were helped but also whether one approach was cheaper (more cost-effective) than another. Other questions may concern whether improvement has been made in a certain staff's productivity since last year or whether the program has reached its intended target population. On some occasions, administrators may want to use evaluation data to help garner public support for human services programs. (The public is much more likely to support tax increases for those programs perceived to be "good" than those thought to be ineffective or poorly run.) Program evaluation can also be used in terms of marketing programs to the public. (As a program manager or agency director, having data showing that 92 percent of your clientele say that they

BOX 1.3 Motivations for Program Evaluation

We want to show:	We want to know:
1. That clients are being helped.	Are clients being helped?
2. That clients are satisfied with our services.	Are clients satisfied with the services received?
3. That the program has an impact on some social problem.	Has the program made any real difference?
4. That a program has worth.	Does the program deserve the amount of money spent on it?
5. That one program or approach is better than another.	Is the new intervention better than the old?
6. That the program needs additional staff or resources.	How do we improve this program?
7. That staff are well utilized.	Do staff make efficient use of their time?

would refer their friends or family members to your agency could be very useful information to have on hand.)

Social and human services programs have evolved to combat such social problems as drug abuse. Think for a moment of other social problems in this country. We could begin listing such problems as:

Poverty
Substance abuse
Homelessness
Adolescent pregnancies
Unemployment
Mental illness
Child abuse
Illiteracy
Domestic violence
High infant mortality rates
Crime
Hunger
AIDS

For each social problem, there are hundreds of potential programs. Some of these programs work and need to be continued; others are ineffective. If it cannot be demonstrated that certain programs have any impact on these problems, then further evaluative research may be undertaken to discover why the programs were not successful. There may be very logical reasons; for example, the programs could be poorly managed, underfunded, or poorly conceptualized or designed. There are many

other reasons. As human services professionals, we need to be just as interested in the outcomes of national programs as we are in our local programs. Program evaluation is not to be understood as having application only to the agency that employs us.

Although the examples used thus far have helped us to understand the need for program evaluation primarily at the local level, an immense need remains for program evaluation of national expenditures and programs. For instance, an article in *Brandweek* (April 1998) entitled "Drug Money" has noted that the Partnership for a Drug-Free America and the White House Office of National Drug Control Policy have embarked on an antidrug campaign costing almost $2 billion. The author of the article, Daniel Hill, says that this enormous expenditure of money is backed only by "flimsy" research. Two of the three studies supposedly showing the effectiveness of media antidrug messages had yet to be published, and the author of the third acknowledged that her respondents might have been saying what they thought the researchers wanted to hear.

Several experts have pointed out that no well-controlled studies show that media campaigns are effective in changing behavior. The deputy director of the Office of National Drug Control Policy was, according to the Hill article, unable to cite any research supporting the contention that antidrug advertising works. Although there is no doubt that a serious drug abuse problem exists in this country, should billions of dollars be spent on untested interventions? Shouldn't research support the effectiveness of interventions (even those perceived to be harmless) before vast sums of money are spent on them?

The example of the antidrug advertising campaign is not an out-of-the ordinary one. According to an article in the *New York Times*, in the 1960s and 1970s, the federal government invested billions of dollars on job training "without a clue about what worked and what did not" (Passell, 1993). As a society we need to test new ideas to combat old problems. For example, do monetary incentives to mothers on welfare for using Norplant contraceptives significantly affect the number of children they have? Is offering full college scholarships to low-income students who remain in school, pass their courses, remain drug-free, and do not become pregnant or get in trouble with the law a realistic way to combat poverty? Sometimes the costs of programs are truly staggering. The widely known and popular program called *Head Start* is intended to give educationally and economically disadvantaged children a leg-up on the early years of schooling. Since 1965, taxpayers have paid over $180 billion to provide this program. In 2012, the federal government released its most comprehensive evaluation of the *Head Start* program. The results? "There was little evidence of systematic differences in children's elementary school experiences through third grade, between children provided access to *Head Start* and their counterparts in the control group (see **http://www.acf.hhs. gov/sites /default/files/opre/head_start_executive_summary.pdf**, p. iv). This was a very large scale and well-controlled evaluation and the evidence is now both clear and compelling. Educationally and developmentally, the best evidence does not show that *Head Start* works as intended. As a form of day care, it has enabled parents to remain in the workforce, but the educational services are of little apparent value with early education gains being lost by the third grade, leaving *Head Start* children as educationally and developmentally as behind as their counterparts who did not receive *Head Start* Services.

Whether at the local, state, or national level, program evaluation often begins by identifying a problem. Decision makers want to distinguish programs that work

from those that do not and to know if their money is well spent. They may have developed questions about a program because of some incident or problems brought to their attention. These problems can be visible and well-recognized or those known only to a handful of staff, administrators, or trustees.

A problem is any undesirable situation or condition. Sometimes program evaluations are undertaken in order to determine the extent or magnitude of a problem or to confirm a suspected problem. As you think about an agency where you have worked or interned, what problems come to mind? (If you do not initially think of any problems, have you seen any recent data suggesting that the program is effective or efficient?)

There are probably as many reasons for conducting program evaluation as there are different programs. In addition to the reasons already given, those in the helping professions also conduct program evaluations because they have a responsibility to improve programs.

We all have an ethical obligation to evaluate our practice. Box 1.4 contains statements taken from the Codes of Ethics of social workers, psychologists, counselors, nurses, medicine, and public health workers, all of which stress the importance of conducting program evaluation (see Box 1.4). All too often, we get caught up in service delivery as measured by billable hours, home visits, numbers of phone calls, and internal audits of agency and accreditation forms without systematically appraising whether all this effort produces beneficial outcomes for clients. We

BOX 1.4 Selected Professional Association Guidelines on the Ethical Mandate to Evaluate Programs and Services

- Social workers should monitor and evaluate policies, the implementation of programs, and practice interventions. Social workers should promote and facilitate evaluation and research to contribute to the development of knowledge. Social workers should … fully use evaluation and research evidence in their professional practice….” (National Association of Social Workers, 1999, p. 20)
- All school social work programs, new or long-standing, should be evaluated on an ongoing basis to determine their relevance, effectiveness, efficiency, and contributions to the process of educating children. (National Association of Social Workers, 1992, p. 16)
- Clinical social workers shall have … knowledge about and skills in using research to evaluate the effectiveness of a service. (National Association of Social Workers, 1989, p. 7)
- There are periodic, systematic, and effective evaluations of psychological services…. When the psychological service unit is a component of a larger organization, regular assessment of progress in achieving goals is provided in the service delivery plan. Such evaluation could include consideration of the effectiveness of psychological services…. (Board of Professional Affairs, Committee on Professional Standards, 1987, p. 719)
- Monitor effectiveness. Counselors continually monitor their effectiveness as professionals and take steps to improve when necessary. (American Counseling Association, 1999, p. 6)

(*continued*)

BOX **1.4** (*continued*)

- The nursing profession should engage in scholarly inquiry to identify, evaluate, refine, and expand the body of knowledge that forms the foundation of its discipline and practice. (American Nursing Association, Code of Ethics, Standard 7.3, available at: **http://nursingworld.org/MainMenuCategories/ThePracticeofProfessionalNursing/EthicsStandards/CodeofEthics.aspx**
- Public health should seek the information needed to implement effective policies and programs that protect and promote health (The importance of information to evaluate programs is also implied).... Evaluate effectiveness, accessibility, and quality of personal and population-based health services (from the Principles of the Ethical Practice of Public Health, version 2.2, Public Health Leadership Society, **xhttp://phls.org/CMSuploads/Principles-of-the-Ethical-Practice-of-PH-Version-2.2-68496.pdf**
- The primary purpose of medical research involving human subjects is to understand the causes, development and effects of diseases and improve preventive, diagnostic and therapeutic interventions (methods, procedures, and treatments). Even the best proven interventions must be evaluated continually through research for their safety, effectiveness, efficiency, accessibility, and quality (from the Declaration of Helsinki of the World Medical Association, **http://www.wma.net/en/30publications/10policies/b3/**)

have an ethical mandate to determine whether our clients are being helped, whether they are any better off as a result of our interventions. Program evaluation is a major means by which we can fulfill this ethical responsibility.

OVERCOMING THE SUBJECTIVE PERSPECTIVE

Anytime we have a choice, we find ourselves in a position where a decision must be made between two or more alternatives. Often, informal (and perhaps even unconscious) criteria guide us in making choices. Although these criteria may be more the product of visceral reactions than of contemplation, they aid us in the making of choices. They help us to determine such things as "good" restaurants and "good" movies, and to rate the services of care providers (e.g., a "good" physician). In each of these instances, good is defined subjectively and somewhat arbitrarily. For example, my notion of the best restaurant in town may be one that specializes in Italian food. You, on the other hand, may intensely dislike Italian cooking. My notion of a good movie may be *Texas Chainsaw Massacre*, whereas your taste may run to less violence. My notion of a good physician may be one who, although known for a disheveled appearance, answers my every question, while your opinion of a good physician requires that the physician dress appropriately and look distinguished. Because appearance is important to you, you may have no confidence in a physician who does not look the role (whether or not your questions get answered).

What does this have to do with program evaluation? Just this: everyday (sometimes many times a day) human services professionals must direct people to their programs or refer them to other programs based on their subjective impressions. When

we make referrals, we want clients to go, not to programs that are not effective, but to the "good" programs. We want them to have the best possible chance of succeeding or doing well in that program. We have a professional responsibility to avoid making referrals to ineffective or deficient programs. We also want the programs we direct or that employ us to benefit our clients. But, how do we recognize a good program, or a poor program?

How do we know when our programs are effective? We like to believe that we help our clients, but what actual evidence do we have that the majority of our clients are helped by our programs? Most helping professionals have had clients who have made giant strides as a result of skilled intervention. We feel rewarded by these successful clients. They help us feel that we are competent and that we have chosen the right career. Unfortunately, there are also those clients with whom we are unsuccessful. These clients, despite our best efforts, drop out of programs, make a mess of their lives, or seem to have gained nothing from our interventions. Think of all the clients who have made their exits from your programs. What is the proportion of "successful" clients to "unsuccessful" clients? Are you more successful than unsuccessful with your clients? What evidence could you present of your success?

We have raised these questions to help you understand that program evaluation involves a different perspective than you may normally employ when thinking about your clients. Clinicians and practitioners tend to evaluate their practice subjectively and in terms of selected individual cases. They think of Mrs. Smith (with whom they were successful), Mr. Arthur (who was a model client and who now comes back to volunteer his services), or perhaps Kathy M., with whom they were not a success. However, this "case focus" does not facilitate the aggregation of data at a program level so that an overall determination can be made about the effectiveness of the program as a whole. Although one bad apple may spoil an entire bushel, one client who does not succeed does not mean that a whole program needs to be overhauled.

The difficulties with attempting evaluation using a "case focus" with a single client can be demonstrated easily. Consider Mrs. Smith. Although you felt that you were successful in helping Mrs. Smith to quit drinking, others may not be so quick to shower accolades on you. Those who are skeptical of your abilities as a clinician may point out that while Mrs. Smith may no longer drink, the rest of her family is in turmoil. Her husband left home; a teenage daughter ran away. Mrs. Smith is now living with another recovering alcoholic and working for minimum wage as a waitress, although she was previously employed as a registered nurse. You reply to these critics, "She's not drinking. She feels good about herself. I think she's shown great improvement." Although it may be possible to argue that any given case was or was not a success, a manager needs to look at the program as a whole. Are the clients (as an aggregate) better or worse off as a result of participating in the program?

Consider the case of Mr. Arthur. Everyone in the agency agrees that he has made significant changes in his life since becoming a client of your program. However, on closer inspection, it is revealed that you spent twice as much time with Mr. Arthur as you did with the average client. Was he a success because he got twice as much attention? Would he have been a success if he had received only as much time as the "average client" receives? (Did he get so much time because he was an "easy" client to work with?)

We have already admitted that the program was not successful with Kathy M. However, is Kathy the typical client or the unusual client? Perhaps Kathy was the most severely disturbed client that your program has ever admitted. Given her previous history of multiple hospitalizations, perhaps no one really expected her to make any significant gains.

We can see from these examples that our perspective as practitioners often involves subjective evaluations. That is, we believe that a client has improved or not improved. The problem with subjective evaluations is that others may not share them. While you think of Mrs. Smith as an example of a successful client, perhaps your best friend and coworker thinks of Mrs. Smith as something less than a success. Although you are quite pleased that Mr. Arthur has overcome a great many of his problems, perhaps your program director has sent you a strongly worded memorandum suggesting that the program's waiting list is such that you are not to spend as much time with the rest of your clients. Although Kathy M. made no progress in treatment, the same program director is not disappointed. "We learned something," she says. "We learned what won't work with clients like this. Next time, we'll try something a little different."

Many people are convinced that they encounter more red lights while driving when they are late, than on days when they have more time. It is highly unlikely the universe is conspiring against them on these hurried days, and much more likely to be an incorrect, but powerful, subjective impression. When surveyed, many psychotherapists rate their overall effectiveness as very high. Like the children of Lake Wobegon, no one appears to be merely "average." Statistically however, this is not possible. In conversation, we can get away with saying things like "I did a good job with that family," "She's a good therapist," or "It's a good program, you'll like it there." However, a thesaurus lists nine different meanings for the word *good* as an adjective (see Box 1.5). Seldom does anyone ask how we define *good*. What we are allowed to do as conversationalists we cannot do as program evaluators.

Evaluators are concerned with specificity and measurements. We want verifiable evidence, not someone's opinion. It matters whether a program produces

BOX 1.5 Denotations of the Word *GOOD*

Usage	Example
1. Pleasant, fine	I had a *good* meal.
2. Moral, virtuous	Mother Theresa was a *good* person.
3. Competent, skilled	She is a *good* worker.
4. Useful, adequate	It was *good* that I read the book before the quiz.
5. Reliable	Pat is a *good* source of information.
6. Kind, giving	My grandmother is so *good*.
7. Authentic, real	He makes a *good* point.
8. Well-behaved	Rachel is such a *good* child.
9. Considerable	There is a *good* deal more poverty now than five years ago.

changes in behavior, attitudes, or knowledge. Further, we might want to know how much change was experienced by the average client, how long it was sustained, and at what cost.

Subjective evaluations about the success of individual clients are very much like the initial examples of a "good" movie and a "good" restaurant. We can expect differences in opinion. Within most groups, if someone says, "That is not a good restaurant!" there are sure to be others within the crowd who will disagree. Someone else may say, "Well, it is my favorite restaurant!" or "That's interesting. We were just there on Wednesday and had a wonderful meal." The problem with subjective evaluations is that everyone is usually right. The person who had a bad experience with a restaurant probably got poor service or an improperly prepared meal. The person who ate there on Wednesday could have just as easily not had a wonderful meal. The individual who boldly proclaimed the restaurant to be his favorite restaurant might be quite willing to forget an occasional bad meal because he goes there for the atmosphere, he is personal friends with the proprietor, or his girlfriend works there. Another possibility is that he just does not have a discriminating palate.

To become evaluators, we need to adjust our perspectives so that we are able to see beyond a single meal or a single client. We need to see the larger picture. We need to go from a micro focus to a macro focus. What are the experiences that most of the restaurant patrons or clients have? In a sense, we need to forget the individual and broaden our perspective to focus on the most common or frequent experience. What percent of the patrons would not return? With what percent of our caseload are we successful? We need to look for corroborative evidence that might convince neutral observers. (For instance, counting the number of patrons leaving meals unfinished or leaving in the middle of a movie might substantiate rather powerfully one's own subjective experience.)

As evaluators, we want to be able to objectively conclude that this program is a good one and that another is not—based not on personal opinion but on factual evidence. When we go beyond our own personal experience or opinions and collect information about the experiences that others have had, we have begun to develop an evaluative stance—we have moved from subjectivity to objectivity.

An objective stance tends to place more reliance in numbers and counting. As a rough rule of thumb, the more individuals we are able to interview, survey, or contact, the more confidence we can place in our evaluative findings. Numbers tend to constitute more objective data (although even statistics can lie!). When, for instance, 97 out of 100 clients indicate that they would recommend our services to their friends, this constitutes objective data. Anyone examining the responses of the 100 clients and sorting them into piles of "would recommend" and "would not recommend" services ought to arrive at the same conclusion.

Evaluators are, in some respects, applied scientists. Scientists seek to understand and explain the world around them. However, it is not just explanations that scientists seek, but *correct* explanations. Whether we think of ourselves as program evaluators or as applied scientists, our findings must stand independently, apart from our claims or persuasive oratory. Our findings should be replicable (reproducible); others must be able to independently arrive at the same conclusions. If someone did not like or agree with the findings from a particular program evaluation, then this

person could repeat the evaluation using the same methodology. Assuming that no major changes occurred within the agency in the interim and that the original evaluation methodology was sound, findings from the second study should be the same or very similar to those of the first study.

Objectivity demands precision. Evaluators must be precise about the program they are evaluating, what they will be measuring, how they will collect and analyze their data, and who they will be interviewing or observing during a given time period. Such matters require specificity. Vagueness is rarely tolerated in research or evaluation. Note the lack of specificity in the following: "This evaluation will determine if specialized in-service training on the use of empathy helps nurses perform their jobs better." Do you find it too vague? The statement is unclear because we are left wondering: What nurses are being discussed? Has it been established that empathy is necessary to perform their jobs? What jobs are under consideration? What does it mean to perform better? How is empathy to be measured?

One way that evaluators become more specific and precise is by using operational definitions. An **operational definition** is the way a variable or concept (such as empathy) is to be defined and measured for the purposes of the evaluation. The evaluator may use a standardized scale to measure level of empathy. Or the evaluator may use some sort of behavioral measures, such as the number of times during a therapeutic session the counselor nods affirmatively or makes supportive statements such as "I understand." Counselors may be operationally defined as those holding certain degrees or as all persons who work in a certain program regardless of their educational background.

As one begins to operationally define the key concepts for a proposed evaluation or study, often the vagueness disappears. In the case of the vague statement, "This evaluation will determine if specialized in-service training on the use of empathy helps social workers perform their jobs better," operationally defining important concepts might change it to: "Do social workers with higher levels of empathy place more children in adoptive homes per year than social workers with lower levels of empathy?"

The effort to become more precise does not rule out the subjective experience in program evaluation. Qualitative research in particular can be very useful in attempting to understand client's inner perspectives on the program services they received and of their service providers. Although a single "bad" subjective experience cannot constitute a program evaluation, it may lead to a formal evaluation as a program manager, agency director, or members of the board of directors become concerned about whether an incident or experience reflects what is "really going on" with the program. The program evaluator seeks to understand the "reality" or "truth" about a program. In the process, the evaluator may collect a large number of subjective opinions about the program. Objective evaluations do not rely on the opinions of any one person (no matter how influential), but instead attempt to gain a comprehensive view from the opinions of the aggregate or group.

Because the reality about a program's performance can sometimes be painful and have far-reaching implications (e.g., loss of funds and the corresponding laying off of a number of an agency's employees), program evaluators often seek the best possible objective evidence that they can obtain (given such pragmatic constraints as budget, time, access to clients or their records, and cooperation of the staff).

Having objective or "hard data" to guide decisions about programs is superior to decision making without program evaluation data. By way of analogy, if you were on trial for an offense that you did not commit, you would want your lawyer to present as much objective evidence on your behalf as possible to assist the jury in realizing your innocence. You probably would not feel comfortable in allowing your attorney to hinge the entire case on the subjective testimony of a single character witness who would testify that you were a "good" student or a "good" friend.

This emphasis on overcoming the subjective perspective is aspirational, not necessarily perfectly accomplished. We are all human beings with our own biases, some of which we may be unaware of, and the introduction of this human element is something we strive to reduce to the irreducible minimum so that our conclusions are not significantly compromised. Good science is not a question of conducting a "purely" objective piece of research, which, lacking perfect objectivity, is fatally flawed and its results no better than "other ways of knowing." A more realistic perspective is addressed by Gorenstein (1986, p. 589):

> It makes no sense to reject the potential scientific import of a construct simply because social values may have played some role in its formulation. The question of whether a construct has any scientific importance is an empirical one. It has to do with whether the construct exhibits lawful properties.

PHILOSOPHICAL ASSUMPTIONS OF PROGRAM EVALUATION

As a specialized form of social and behavioral science research, program evaluation is predicated on certain philosophical assumptions pertaining to the nature of reality, and of the design and conduct of research inquiry. Although these philosophical assumptions are often confused with *theory*, you can differentiate them by recalling that social and behavioral science theories are *explanations* of problems or interventions, explanations amenable to investigation via scientific research, and that theories themselves are based on certain philosophical assumptions that are fundamentally untestable. Some of these philosophical foundations of program evaluation are listed in Box 1.6.

Each of these points of view has occupied central controversies within philosophy for hundreds of years, and in some cases, millennia. In offering them as characterizing the philosophy of science undergirding program evaluation, we make no pretense that any one of them can be considered irrefutably justified either by logic or by empirical data. Consider them, like Euclid's axioms, as undemonstrated propositions (although common sense would suggest that they are self-evident) that serve as intellectual pivot points around which program evaluation efforts revolve.

No one, for example—certainly not the authors—is capable of providing a philosophically irrefutable proof that the world has a physical reality (as opposed to a subjective construction in one's mind). But in order to take action, some assumptions are necessary. Thus, for example, the distinguished social work educator Bertha Capen Reynolds claimed, "A second characteristic of scientifically oriented social work is that it accepts the objective reality of forces outside itself with which it must cooperate" (Reynolds, 1942, p. 24). Can this assumption of an

BOX 1.6 Selected Philosophical Foundations of Program Evaluation

Acceptance of:

Realism—The point of view that the world has an independent or objective existence apart from the perceptions of the observer

Determinism—The assumption that all phenomena have physical causes that are potentially amenable to scientific investigation

Positivism—The belief that valid knowledge about the objective world can be arrived at through scientific research

Rationalism—The belief that reason and logic are useful tools for scientific inquiry, and that, ultimately, truthful explanations of human behavior will be rational

Empiricism—A preference to rely on evidence gathered systematically through observation or experiment, and capable of being replicated (i.e., verified) by others, using satisfactory standards of evidence

Operationalism—The assertion that it is important to develop measures of phenomena (e.g., client problems, interventive procedures) that can be reliably replicated by others

Parsimony—A preference for the simpler of the available adequate explanations for behavioral phenomena

Pragmatism—The view that the meaning or truth of anything resides in its consequences in action

Scientific skepticism—The point of view that all claims should be considered of doubtful validity, until substantiated by credible scientific data

Rejection of:

Nihilism—A doctrine that all values are baseless and that nothing is knowable or can be communicated

Anecdotalism—The belief that anecdotes prove something empirically

Metaphysics—Explanation involving supernatural, incorporeal or immaterial entities or factors

Dualism—The view that the world consists of the two fundamental entities of mind and matter

Reification—Attributing reality status to an abstract or hypothetical construct (e.g., the superego) in the absence of adequate evidence supporting the existence of that construct

Circular reasoning—An explanation for human behavior in which causes and effects cannot be distinguished from each other

Scientism—The theory that the investigational methods used in the natural sciences should be applied to answer *all* questions of concern to people (e.g., ethical questions, moral questions, etc.)

Solipsism—The view that there is no proof that anything exists outside the mind

objective reality, known in philosophy as **realism**, be irrefutably shown to be true? No, but for the purposes of attempting to improve the condition of the world we live in, some such practical assumptions must be accepted.

We urge the reader to understand that the assumption of the validity of some of these philosophical beliefs does not preclude accepting others, even those apparently at odds with the former. For example, the assumption of realism does not disavow the importance of individual clients' perceptions of their world, or of the meaning they ascribe to their lives and relationships. Nor does it deny that to some extent individuals construct their own reality. But realism asserts that there is an element to the client's world that objectively exists, and that this is often a very important one, in fact often the *most* important to program evaluators. For example, evaluations of domestic violence interventions are more likely to concentrate on ascertaining whether or not acts of violence *really* declined or ceased, as opposed to trying to understand a spouse's "meaning" of what it means to be beaten. The evaluation of a neighborhood watch crime reduction program may look at community members' perceptions of crime rates, but the *actual* rates of various crimes before and after the implementation of the neighborhood watch program are usually seen as more important. Program evaluations usually focus on determining objective *changes* in the lives of clients, not on understanding subjective processes. Of course, it is a legitimate part of science to study such subjective processes, such as the meaning to clients of changes brought about by a program, but such is not usually the primary focus of program evaluation. Sometimes evaluation studies encompass both aspects of the impact of programs—objective and subjective—and that can be worthwhile.

To accept **determinism** does not deny that many other phenomena (e.g., psychosocial problems) exist and possess such a complex constellation of interlocking etiological factors that creates an appearance of chaos or randomness at times. And it may be that only a portion of a particular psychosocial phenomena is "determined" in a scientific sense. If so, then it is the task of scientific inquiry and of program evaluation to fully explore the limits of this determinism, even if it yields an understanding less than 100 percent complete. A "Program of Assertive Community Treatment" that helps reduce psychiatric rehospitalization by 30 percent among the chronically mentally ill is not a failure because it is not 100 percent effective. Clearly, a 30 percent reduction leaves much to be understood and plenty of room for additional improvements. But 30 percent is better than where the clients were previously. To be a **positivist** is not to assert that certain knowledge in a particular area currently exists, but rather to claim that it is a good idea to strive for such knowledge through scientific methods because this may result in ever closer approximations to truth. This is called scientific progress. Although one may be an **empiricist**, this need not deny the role of common sense, intuition, authority, and tradition as potentially valuable sources of knowledge.

Rationalism contends that factual explanations for psychosocial phenomena can be arrived at via logic—armchair reasoning if you will. Unfortunately, rationalism can also lead to errors. Thus, rationalism and logical thinking are seen as components of ultimately truthful explanations (e.g., being able to explain logically why a program works), but these conclusions *must* be evaluated using empirical tests, as these are the best means of sorting out truth from fiction.

Parsimony in theory is nothing new. Clinicians are taught, for example, to exclude organic factors as possible causes of a client's depression prior to implementing an intervention based on a presumptive psychodynamic etiology. Similarly, a theoretical explanation that uses fewer unverified factors is generally preferred over one that invokes more complex accounts. Consider, for instance, that for many years homosexual orientation was associated with severe psychopathology. Why? Well, many gay and lesbian people sought treatment from psychiatrists and other mental health professionals. Thus, they appeared to have greater pathology than the general population. Was the assumption correct? One possibility was that homosexual orientation *was* associated with mental disorder. The second possibility was that this conclusion was only a distortion caused by a selection bias (those seeking treatment for mental and emotional problems were not representative of all gays and lesbians). Parsimony suggested that careful attention be given to the latter explanation prior to accepting the former; and indeed, studies have found that gays and lesbians are no more or less mentally disordered than heterosexuals (Thyer, 2014). Certainly not all parsimonious explanations possess greater validity than more complex reasons. After all, many psychosocial problems are *really* very complex. Parsimony simply cautions us to rule out or entertain potentially simpler accounts prior to accepting more involved ones.

Pragmatism, although commonly seen as synonymous with practicality, has a more in-depth meaning involving the *consequences* of some program. According to pragmatism, the "meaning" of a child abuse prevention program resides in the numbers of children who are not exposed to abuse as a result of that program. A job-finding program's real meaning is based on the number of clients successfully placed in good jobs. A program or intervention that fails to produce any changes can be said, in a pragmatic sense, to be a meaningless program. The meaning of a program is not how clients or service providers view the service, but stems from the practical, positive results obtained because of involvement with it. Pragmatism does not exclude other (perhaps more subjectivist) interpretations of what a program may mean to individuals; but for the purposes of program evaluation and practice in the human services, it refers to the practical outcomes of an intervention. And that, after all, is what program evaluation is aimed at finding out.

Scientific skepticism is the profession's shield against the premature adoption of unsupported claims. The burden of proof lays at the feet of the person making an unusual claim. "Facilitated communication helps kids with an autistic disorder communicate." "Neurolinguistic programming is an effective treatment for phobias." "Primal scream therapy helps people who are depressed." It is not difficult for the reader to encounter such claims every day. Scientific skepticism is the modest request that such assertions be supported with credible evidence prior to being accepted or widely adopted in practice. The burden of proof is not on the skeptic to show that these things do not work—it rests with the proponents to demonstrate that they do. Unlike the doctrine of **nihilism**, a point of view fatal to the spirit of scientific inquiry, skepticism is a tempered perspective. It does not deny the possibility that genuinely truthful explanations can be obtained; skepticism requires only that appropriate proofs be provided for positive assertions. Whenever a human services professional encounters a proposal to fund a new program, it is always a good idea (scientifically, if not politically) to ask, "Where are the data to support this approach?"

We like the idea of using the national census to illustrate these principles. At a given point in time, *one* number represents how many persons live in the United States (realism). The census is designed to try and capture this mythic number as closely as possible (positivism). The scientific methods used by the Bureau of Census attempt to use empirical and operationalized methods to obtain data from U.S. residents. This is certainly not a perfect approach to trying to gain an accurate census—minorities of color, undocumented aliens, and other marginalized persons (e.g., the homeless) are undercounted. However, no reasonable alternative methodologies outside of conventional scientific inquiry can be employed for this purpose. We cannot simply ask some noted authority, "Excuse me, Mr. President, how many people are there in the United States at this moment?" Our intuition or even personal practice experience cannot help, and it is extremely unlikely that divine revelation will lend a hand. Similarly, it can be assumed that certain outcomes follow clients' experiences with a particular human services program, and it is the task of the program evaluator to best ascertain what these outcomes may be. Cumbersome and imperfect though they may sometimes be, there is no substitute for systematic, empirically oriented efforts at evaluation guided by the methods and philosophy of mainstream science, which, parenthetically, can embrace both quantitative and qualitative research approaches.

Lastly, the notion of what might be called **anecdotalism**—the belief that anecdotes prove something empirically—needs to be addressed. Personal experiences and accounts can create very powerful belief systems that may yield true conclusions in a situation or for a particular individual. Someone who is mugged by a bearded bandit may justifiably be afraid of this fellow the next time he is encountered; however, to generalize that fear to all bearded individuals would be unfortunate. The problem with anecdotes is that their "truth" may not generalize. Not all bearded people are threats.

At the same time, the personal experiences of clients and staff with a program are important to the program evaluator. We want to know about the problems encountered and the good things that resulted. These experiences may become part of a qualitative evaluation or aggregated into more of a quantitative evaluation. However, we should not assume that an individual anecdote (or even several of them) constitutes a *philosophical foundation* for program evaluation. Although a single event (e.g., death of a client) may trigger an evaluation, that event does not provide a worldview guiding all inquiry as do the positions previously discussed (scientific skepticism, pragmatism, etc.). Science does not reject anecdotes as false, or claim that all are false, but merely asserts that as evidence they provide a low level of proof and that they are most meaningful when augmented by more robust standards of evidence. It has been justifiably said that the plural of anecdote is not data.

MORE ON POSITIVISM

Much discussion has been expended in the program evaluation literatures over the meaning of the term *positivism*, and of its value as a foundation for inquiry in the human services (Bolland & Atherton, 2002; Smith, 2002; Thyer, 1993, 2008b, 2008c). Keep in mind that from its inception, professional social work accepted

the doctrine of positivism, established in the early part of the 1800s by the French philosopher and scientist Auguste Comte. Comte is said to have established the discipline of sociology, which he originally called *social physics*, based on his contention that human behavior could be studied using the same tools and principles that science used to study natural phenomena:

> The word *positive* came from *ponere* and had been employed since the fourteenth century to mean *laid down*. In the sixteenth century, it began to refer to knowledge that was based on facts and was thus reasonably certain. Eighteenth century thinkers used the word *positive* to oppose the *metaphysical*. (Pickering, 1993, p. 65)

According to Comte, "Unlike theological beliefs, scientific truths could be proved if necessary.... [S]cientific *truths* were always provisional because they could be proved wrong at any time" (Pickering, 1993, p. 171). Despite its name, positivism took a provisional approach to the development of knowledge, with Comte noting in 1825 that "Scientific laws were only *hypotheses* constructed by man with *external materials* and confirmed by observation; they amounted to no more than approximations of a reality that could never be rigorously understood" (Pickering, 1993, p. 294).

From its inception, positivism was concerned not just with studying human phenomena, but also with improving the human condition. "Social science was not ... just an intellectual mixture of history, the physical sciences, physiology, and political economy. It had a practical vocation: *to regenerate society*.... Although Comte admitted that we could never fully know external reality, he assumed that scientific theories were getting closer to representing it *exactly*" (Pickering, 1993, pp. 294, 296, italics added). Positivism came to exert an enormous influence on science, on the emerging social sciences, and in social welfare in particular. In the United States, the American Social Science Association (ASSA) was established in 1865, and set forth as its mandate:

> *Social science* was understood by (ASSA) members to refer to the whole realm of problematical relationships in human affairs. One became a social scientist "by contributing to the store of esoteric knowledge and practical expertises ... *a new way to care for the insane or to administer charity*—all of these were equally valuable contributions to *social science*." (Haskell, 1997, pp. 97–98, italics added)
>
> This Association proposes to afford to all persons interested in human improvement, an opportunity to consider social economics as a whole.... They are to collect all facts, diffuse all knowledge, and stimulate all inquiry, which have a bearing on *social welfare*. (Haskell, 1997, p. 102, italics added)

From the ASSA emerged, in 1879, the Conference on Charities, transformed in 1884 into the National Conference of Charities and Corrections (NCCC). In 1917 the NCCC became the National Conference on Social Work, transformed again into the National Conference on Social Welfare in 1957, which in turn dissolved in the mid-1980s. A paper presented at the 1889 meeting of the NCCC was titled "Scientific Charity" and an article appearing in an 1894 issue of the influential journal *The Charities Review* was titled "A Scientific Basis for Charity." Such early works were a part of the movement called scientific charity (or scientific philanthropy), which had its own origins in the 1870s, again based on the fundamental assumptions of positivism.

The distinguished social work educator Frank Bruno (1964) provided this overview of a National Conference on Charities meeting of the late 1800s:

> Most of the leaders of the Conference accepted the implications of a scientific approach to social work problems. They acted on the tacit assumption that human ills—sickness, insanity, crime, poverty—could be subjected to study and methods of treatment, and that a theory of prevention could be formulated as well.... This attitude raised these problems out of the realm of mysticism into that of a science.... As a result of the adoption of this scientific attitude, Conference speakers and programs looked toward progress, not backward toward a golden age. They believed in the future; that it was possible, by patient, careful study and experimentation, to create a society much better than the one they lived in. (pp. 26–27)

And about 30 years earlier Bruno (1936) had provided his own congruent perspective on the integration of science and social work:

> Social work holds as its primary axiom that knowledge of human behavior can be acquired and interpreted by the senses and that inferences drawn from such knowledge can be tested by the principles of logic. The difference between the social work of the present and all of the preceding ages is the assumption that human behavior can be understood and is determined by causes which can be explained. We may not at present have a mastery of the methods of understanding behavior, but any scientific approach to behavior presupposed that it is not in its nature incomprehensible by sensory perceptions and inference therefrom. It follows from such a theory that understanding is the first step in the direction of control and that the various forms of human misery are susceptible not only of amelioration, which our predecessors attempted, but also of prevention or even of elimination, when once their nature is understood. (pp. 192–193)

Both the statements, though they may be dated, could be said to characterize the mainstream contemporary perspective adhered to by science in general, and by program evaluation in particular. A more recent synopsis of positivism was put forth by Glashow (1989, p. 24E):

> We believe that the world is knowable, that there are simple rules governing the behavior of matter and the evolution of the universe. We affirm that there are eternal, objective, extra historical, socially neural, external and universal truths and that the assemblage of these truths is what we call ... science. Natural laws can be discovered that are universal, invariable, inviolate, genderless and verifiable. They may be found by men or women ...

You will encounter the term *positivism* in your academic studies, and in the program evaluation literature. Keep in mind the relatively simple premise behind it: "A paradigm introduced by Auguste Comte, which held that social behavior could be studied and understood in a rational, scientific manner—in contrast to explanations based on religion or superstition" (Rubin & Babbie, 2008, p. 642). Positivism is an approach, not intended to represent accomplished facts about human behavior. As Todd noted back in 1920, "Science does not claim to have complete knowledge for the world or to have established perfect order out of chaos in this world. It is less an accomplished fact than an attitude" (p. 71). Nor does positivism deny the potential utility of other nonscientific approaches to knowledge development in social work—tradition, values and ethics, religion, mysticism; all may have a valuable contributing role to our field. Asserting that one

approach (called A) is useful, is not to say that B is valueless. Positivism does make strong claims regarding the value of scientific approaches to knowledge development in general and program evaluation in particular. Some may even assert it makes the strongest claims. What is important is the development of reliable and valid information, and this can be approached through many methods, including positivistic ones.

CHAPTER RECAP

Whether you are a direct service worker, program director, an agency administrator, or a policy-maker, you want the organization that employs you to be well managed and responsive to the needs of clients and community. How does an agency become a well-managed agency? One essential way is the evaluation of its efforts, where problems are identified and corrective action taken.

What is essential to learn about program evaluation? Besides understanding the purpose of program evaluations and some of the various reasons why they are conducted, you need to know the difference between a subjectively held opinion and one that is derived from objective data. This book will help you develop ways of identifying, collecting, and using data that will allow you to be as objective as possible when evaluating programs in the social and human services. Objective data are seen as having greater credibility and as providing better information for the decisions that face program managers. Evaluators use operational definitions to obtain objective data that can be replicated if necessary. Theoretical models suggest not only what interventions might work, but also where to find the changes that have resulted.

Questions for Class Discussion

1. Make a list of two or three human services programs with which you or members of the class are familiar. In another column, list what is known about how well each program does its job. For example, what is its success rate? Other than subjective feelings about these programs, what is known about how "good" these programs are? In a third column, make a list of questions that you would like to have answered about each program.
2. Evaluators must operationally define what will be recognized as "success" or a "successful outcome." Try your hand at operationally defining "success" for one of the programs you listed in question 1.
3. For the human services programs you listed in question 1, discuss your ideas about theoretical orientations on which the interventions are, or might be, based.
4. Discuss your experiences, if any, with program evaluation in current or past practice experiences.
5. What are the characteristics of a "good" movie? Make a list of all the subjective opinions held by the class members about a "good" movie. How could you objectively determine if a movie is "good"?

6. Why is it necessary to develop operational definitions about such things as what constitutes recidivism or a successful client outcome?
7. Pick one of the philosophical assumptions undergirding program evaluation described in Box 1.6 and defend it as a useful orientation for program evaluation. Pick another assumption and present some reasoned arguments as to why you do not think it to be a useful perspective.
8. Assume that the "perfect" program evaluation has been conducted (sadly it has not) and it found that participation in Alcoholics Anonymous leads to sobriety. Discuss what implications this has for the theory underlying AA.
9. Distinguish between research-supported treatments, and the practice model of evidence-based practice (EBP). Why, according to the EBP model, are there no such things as evidence-based *practices or treatments*?

Mini-Projects: Experiencing Evaluation Firsthand

1. Choose a product (e.g., coffeemakers, laptop, electronic notebook, cell phone), and develop a set of objective standards that could help consumers select a model of superior performance and avoid the inferior models. Once you have finished, consult back issues of *Consumer Reports* to see how the standards you used compare with those used by the Consumer Products Testing Union.
2. What would you request in the way of an evaluation if you were in a position to require evaluation of a national program in the health or human services fields? Select a national program, and identify what information would be needed in order for an unbiased panel of experts to conclude that the program was successful.
3. Find an example of a program evaluation study published in a professional journal. We suggest that you learn to use the online database called PsycINFO, which is likely available via your local university library, to locate such studies. Briefly describe how key variables were operationally defined in this study, how the program's success was measured, and whether or not the program explicitly made use of an explanatory theory—etiological or interventive, or both.
4. Using one of the websites found in Table 1.2, locate a psychosocial intervention that is said to be research-supported. Describe the evidence, which was used to label this intervention as research- or empirically supported.

References and Resources

American Counseling Association. (1999). *Code of ethics*. Alexandria, VA: Author. Retrieved from www.counseling.org/resources/codeofethics.htm

American Psychiatric Association. (2013). *Diagnostic and statistical manual for mental Disorders-5*. Washington, DC: Author.

Angell, R. C. (1954). A research basis for welfare practice. *Social Work Journal*, *35*, 145–148, 169–171.

Austin, D. M. (1998). *A report on progress in the development of research resources in social work*. Austin, TX: University of Texas School of Social Work.

Barker, R. (Ed.). (2014). *The social work dictionary* (6th ed.). Washington, DC: NASW Press.

Board of Professional Affairs, Committee on Professional Standards. (1987). General guidelines for providers of psychological services. *American Psychologist*, *42*, 712–723.

Bolland, K., & Atherton, C. (2002). Heuristics versus logical positivism: Solving the wrong problem. *Families in Society*, *83*, 7–13.
Bruno, F. (1936). *The theory of social work*. New York: Health.
Bruno, F. (1964). *Trends in social work: 1874–1956*. New York: Columbia University Press.
Cabot, R. (1931). Treatment in social casework and the need of criteria and of tests of its success or failure. In Proceedings of the National Conference of Social Work (pp. 3–24). Chicago: University of Chicago Press.
Chambless, D., Sanderson, W., Shoham, V., Johnson, S., Pope, K., Crits-Cristoph, P., et al. (1996). An update on empirically validated therapies. *The Clinical Psychologist*, *49*(2), 5–18.
Conrad, K. J., & Miller, T. Q. (1987). Measuring and testing program philosophy. *New Directions for Program Evaluation*, *33*, 19–42.
Corsini, R. (2002). *The dictionary of psychology*. New York: Brunner-Routledge.
Ell, K. (1996). Social work research and health care practice and policy: A psychosocial research agenda. *Social Work*, *41*, 583–592.
Ferreira, R. J., & Buttell, F. P. (in press). Can a 'psychosocial model' help explain violence perpetuated by female batterers? *Research on Social Work Practice*.
Glashow, S. (1989, October 22). Positivism lives: We believe that the world is knowable. *New York Times*, p. 24E.
Gorenstein, E. E. (1986). On the distinction between science and valuation in the mental health field [letter]. *American Psychologist*, *41*, 588–590.
Greenwood, E. (1957). Social work research: A decade of reappraisal. *Social Service Review*, *31*, 311–320.
Grinnell, R. M., & Unrau, Y. A. (Eds.). (2008). *Social work research and evaluation* (8th ed.). New York: Oxford University Press.
Haskell, T. L. (1997). *The emergence of professional social science: The American Social Science Association*. Urbana, IL: The University of Illinois Press.
Hibbs, E. D., & Jensen, P. S. (Eds.). (2005). *Psychosocial treatments for child and adolescent disorders: Empirically based strategies for clinical practice* (2nd ed.). Washington, DC: American Psychological Association Press.
Hill, D. (1998). Drug money. *Brandweek*, *39*(17), 20–26.
Holosko, M. (2010). What types of designs are we using in social work research and evaluation? *Research on Social Work Practice*, *20*, 665–673.
Kolata, G. (1994, May 4). Clinic entices patients by paying them $10 a visit. *New York Times*, p. B8.
LeCroy, C. W., & Williams, L. R. (2013). Outcome studies in social work journals: A review of publication trends. *Research on Social Work Practice*, *23*, 706–709.
Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2014). Why ineffective psychotherapies appear to work: A taxonomy of causes of spurious therapeutic effectiveness. *Perspectives on Psychological Science*, *9*, 355–387.
Lohr, J. M., Hooke, J., Gist, R., & Tolin, D. F. (2003). Novel and controversial treatments for trauma-related stress disorders. In S. D. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 243–272). New York: Guilford Press.
MacKenzie, D. L., Biere, D., & Mitchell, O. (2007). An experimental study of a therapeutic boot camp: Impact on impulses, attitudes and recidivism. *Journal of Experimental Criminology*, *3*, 221–241.
Mullen, E. (1995). A review of research utilization in the social services. *Social Work*, *40*, 282–283.
National Association of Social Workers. (1989). *NASW standards for the practice of clinical social work*. Silver Spring, MD: Author.

National Association of Social Workers. (1992). *NASW standards for school social work services*. Washington, DC: NASW Press.
National Association of Social Workers. (1999). *Code of ethics*. Washington, DC: NASW Press.
Passell, P. (1993, March 9). Like a new drug, social programs are put to the test. *New York Times,* p. B5.
Pickering, M. (1993). *Auguste Comte: An intellectual biography* (Vol. I). New York: Cambridge University Press.
Pignotti, M., & Thyer, B. A. (2009). Some comments on *Energy psychology: A review of the evidence:* Premature conclusions based on incomplete evidence? *Psychotherapy: Theory, Research, Practice, Training*, *46*, 257–261.
Reynolds, B. C. (1942). *Learning and teaching in the practice of social work*. New York: Farrar & Rinehart.
Rosen, A., Proctor, E. K., & Staudt, M. M. (1999). Social work research and the quest for effective practice. *Social Work Research*, *23*, 4–14.
Rosenberg, M. L., & Brody, R. (1974). The threat or challenge of accountability. *Social Work*, *19*, 344–350.
Rubin, A., & Babbie, E. (2008). *Research methods for social work* (6th ed.). Belmont, CA: Thomson.
Rutter, M. (1987). Psychosocial resilience and protective mechanisms. *American Journal of Orthopsychiatry*, *57*, 316–331.
Saunders, D. G. (1992). A typology of men who batter: Three types derived from cluster analysis. *American Journal of Orthopsychiatry*, *62*, 264–275.
Schalock, R. L., & Thornton, C. V. D. (1988). *Program evaluation: A field guide for administrators*. New York: Plenum.
Smith, D. (2002). The limits of positivism revisited. *Social Work and Social Sciences Review*, *10*, 27–37.
Stanley, M. A., Wilson, N. L., Amspoker, A. B., Kraus-Schuman, C., Wagener, P. D., Calleo, J. S., et al. (2014). Lay providers can deliver effective cognitive behavior therapy for older adults with Generalized Anxiety Disorder: A randomized trial. *Depression & Anxiety*, *31*, 391–401.
Strauss, S. E., Glasziou, P., Richardson, W. S., & Haynes, B. (2011). *Evidence-based medicine: How to practice and teach EBM* (4th ed.). New York: Elsevier.
Sugarman, B. (1988). The well-managed human service organization: Criteria for a management audit. *Administration in Social Work*, *12*(4), 12–27.
Taubman, S. L., Allen, H. L., Wright, B. J., Baicker, K., & Finkelstein, A. N. (2014). Medicaid increases emergency-department use: Evidence from Oregon's Health Insurance Experiment. *Science*, *343*, 264–268.
Thomson, H., Thomas, S., Sellstrom, E., & Pattecrew, M. (2013). *Housing improvements for health and associated socioeconomic outcomes: A systematic review*. Retrieved from http://www.campbellcollaboration.org/lib/project/61/
Thyer, B. A. (1993). Social work theory and practice research: The approach of logical positivism. *Social Work and Social Sciences Review*, *4*(1), 5–26.
Thyer, B. A. (1994). Are theories for practice necessary? *Journal of Social Work Education*, *30*, 147–151.
Thyer, B. A. (2001). What is the role of theory in research on social work practice? *Journal of Social Work Education*, *37*, 9–25.
Thyer, B. A. (2004). What is evidence-based practice? *Brief Treatment and Crisis Intervention*, *4*, 167–176.
Thyer, B. A. (2008a). Evidence-based practice, science, and social work. In A. Roberts (Ed.), *Social worker's desk reference* (pp. 115–1119) (2nd ed.). New York: Oxford University Press.