1

Evidence-Based Practice

An Introduction

Peter Sturmey

Evidence-Based Practice

Questions concerning evidence-based practice (EBP) permeate services for people with developmental disabilities. A parent must decide whether or not to participate in early intervention for their child with autism and stay at home instead of working. A teacher selects certain target behaviors and teaching strategies for a child and decides not to teach other skills and not to use other teaching strategies. An agency for adults with disabilities decides whether to operate traditional, center-based services or to implement a new job coaching service. A city, state, or government agency decides whether or not to fund early intervention or to place some individuals in specialized, expensive, out-of-district services.

EBP is reflected in many educational and clinical decisions by individual teams. Consider the following example. A team of professionals in special education attempt to treat food refusal in a child with autism for 6 months using sensory integration therapy. Not only did the child continue to refuse food, but the child continued to lose weight. When outside therapists proposed using escape extinction (see Chapter 8) as an EBP, the educators oppose such treatment on the grounds that it does not address the sensory needs of the child and will not work or they refuse to treat the problem by “conditioning” or “behavior modification” which they think is “inhuman” or “disrespectful.” After 2 weeks of escape extinction, the child now ate a wide range of foods and gained weight; resources are no longer wasted on ineffective therapy and useless discussion of ineffective treatment.

EBP is not some academic question. If we are concerned with personally significant outcomes and avoidance of harm for individuals with developmental

Evidence-Based Practice and Intellectual Disabilities, First Edition.
Edited by Peter Sturmey and Robert Didden.

disabilities, it is one that we all face. This chapter outlines some of the issues in the application of EBP to services for people with intellectual disabilities (ID), autism, and other developmental disabilities. The next section examines the general and operational definitions of EBP. The next sections examine the ethical and economic rationales for EBP and the methods associated with EBP, such as systematic reviews and meta-analyses. The final section reviews some of the application of EBP to ID, autism, and other developmental disabilities.

What Is Evidence-Based Practice?

Some definitions

Some general definitions

The definition of EBP is anticipated in Paul's (1967) famous questions: "What treatment, by whom, is most effective for this individual with that specific problem, and which set of circumstances?" (p. 111) which—nearly 50 years ago—raised the issue of not only what kinds of psychological therapy are effective but also how a practitioner should apply or not apply the results of therapy outcome research to specific clients with specific problems. Paul's question is echoed in Sackett, Richardson, Rosenberg, and Haynes' (1997) definition of EBP as "the integration of best research evidence with clinical expertise and patient values" (p. 1). This definition is cited very often and is the basis for similar definitions, such as those by the Institute of Medicine (2001) and the American Psychological Association's (APA) Presidential Task Force on EBP (2006).

Table 1.1 lists a number of definitions of EBP. An examination of these definitions shows that they are aspirational rather than operational, as they do not describe the methods by which we might determine and apply best research evidence clearly. For example, the meaning of the words such as "most effective," "integrate," "clinical

Table 1.1 Some Definitions of EBP

1. "… the integration of best research evidence with clinical expertise and patient values" (Sackett et al., 1997, p. 1)
2. "Evidence-based practice in psychology (EBPP) is the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences" (APA, 2006, p. 273) (http://www.apa.org/practice/resources/evidence/evidence-based-statement.pdf)
3. "… a decision-making process that integrates the best available evidence with family and professional wisdom and values" (Buysse & Wesley, 2006, p. 12)
4. EBP early childhood intervention practices are "informed by research, in which the characteristics and consequences of environmental variables are empirically established and the relationship directly informs what a practitioner can do to produce a desired outcome" (Dunst, Trivette, & Cutspec, 2002, p. 3)

expertise," "wisdom" and "values," and "informed by research" is not specified in these definitions, although some sources, such as Straus, Glasziou, Richardson, and Haynes (2011), do describe specific procedures that practitioners can use to determine what is EBP and how to use it effectively with specific particular clients. Some definitions give greater emphasis to science and EBP, such as Sackett et al.'s (1997) and Dunst et al.'s definitions. Other definitions, such as Buysse and Wesley's, appear to give equal weight to clinical experience and wisdom and only reference "best available evidence" rather than research or empirically validated treatments. Some definitions emphasize that evidence must be current and the "best available evidence" but also that this evidence must be integrated into clinical decision making. Finally, all of these definitions state that evidence and experience must be "integrated," but these definitions do not specify what constitutes integration or how it can be achieved.

These definitions illustrate the tension between the role of science and personal and professional experience should have in determining what an appropriate treatment is for a specific person. They also reflect the tension between personal and professional autonomy and choice over selection of treatment versus restriction of choice and autonomy implied by restriction of practice to only EBP by funding agencies and professional organizations. For example, the explicit aim of some meta-analyses is to determine the standards of practice in developmental disabilities (Scotti, Evans, Meyer, & Walker, 1991). Professional practice is also restricted by treatment algorithms and practice guidelines. For example, National Institute for Clinical Excellence's (NICE) (2012) guidelines for adults with autism baldly state, "Do not provide facilitated communication" (p. 24). They also guide professionals as how to conduct certain treatments. For example, when conducting social learning programs for adults with autism, the guidelines state that they should "typically include: … modeling … peer feedback (for group-based programs) or individual feedback (for individually delivered programs) … discussion and decision-making … explicit rules … [and] … suggested strategies for dealing with difficult social situations." A clinician who does not follow such evidence-based, professional practice guidelines would have to justify deviations from them or use of alternate approaches and might encounter censure during peer review or any legal proceedings if a client or someone else is harmed. A final important observation on these definitions is that they go beyond merely generating lists of treatment that meet the criteria for EBP. EBP also requires the application and adaption of research findings to the actual effective delivery of the best intervention to real-world clients in real-world settings by local practitioners to achieve actual client outcomes for the specific client at hand.

Operational definitions

As well as these general definitions of EBP, there are also operational definitions of EBP. Here, we can discern two kinds of approaches. The first is to determine if a specific treatment is an EBP. For example, Chambless and Hollon (1998) operationally defined an EBP in two ways. When discussing randomized controlled trails (RCTs), they write that "Only when a treatment has been found efficacious in at least two studies by independent research teams do we consider its efficacy to have been established and label it an efficacious treatment. If there is only one study supporting a treatment's efficacy, or if all of the research has been conducted by one team, we consider the findings promising but would label such treatments as possibly efficacious, pending replication" (p. 10), and later, when discussing small *N* experiments, they wrote that "We consider a treatment to be possibly efficacious if it has proved beneficial to at least three participants in research by a single group. Multiple replications (at least three each) by two or more independent research groups are required before we consider a treatment's efficacy as established (each in the absence of conflicting data)."

The second approach is to determine what an EBP is for a specific client and a specific presenting problem. Straus et al. (2011) described a five-step procedure that a practitioner should use to identify an EBP for a specific clinical situation. Step 1 was to convert an unmet information need into an answerable question. Step 2 was to find the best evidence available to answer that question. Step 3 was to appraise the evidence critically as to its validity, effect size (ES), and applicability to the situation at hand. The fourth step was to integrate the results of step 3 with practitioner expertise and the client's biology, values, and circumstances. Finally, step 5 was to implement the EBP with the particular client and evaluate both the effectiveness and efficiency of that intervention. NICE's (2012) description of the six steps of EBP is a modification of these guidelines. This approach to defining EBP is closer to clinical practice and fulfills the function of taking a clinical problem and translating research on treatment outcome to application with a specific client.

These two approaches illustrate different procedures because they reflect two different kinds of questions. The first kind of question is a research question relating to the current status of scientific evidence generally for a treatment, a problem, or the application of a treatment or range of treatments to a problem. For example, we might ask a very broad question such as "What is the most effective treatment for challenging behavior?" or we might ask more focused questions such as "Does functional analysis produce more effective treatments than functional assessment for the treatment of aggression in preschool children with autism?" or "What are the effects of PECS training on requesting, spoken language and challenging behavior?" These questions result in identification of specific treatments as EBPs and may generate lists of approved and nonapproved treatments or treatments rank ordered in various ways. For example, treatments may be ranked by degree of empirical support, ES, risk of harm, cost, etc., or some combination of these. For example, Chambless and Hollon's (1998) criteria classified therapies as "possibly efficacious" or "efficacious." The second kind of question related to how a practitioner can apply the available evidence to a specific client and, thus, is closer to answering Paul's (1967) question. For example, here, we might ask, "What is the best treatment of self-injury for this teenager with Lesch–Nyhan syndrome that we can implement and that he and his family will support?"

Other definitions

Chambless and Hollon's (1998) criteria for EBP are often cited, but they are not the only operational definition of EBP. For example, Chambless and Hollon noted that their own criteria differed from those of the Division 12 task force (1998) which defined EBPs as having *specificity*; that is, for the Division 12 task force (1998), an EBP had to be superior to sugar pill or psychological placebo and another psychological treatment to be considered an EBP. In contrast, Chambless and Hollon only required superiority to a wait list control, arguing that "if a treatment works, for whatever reason, and if this effect can be replicated by multiple independent groups, then the treatment is likely to be of value clinically, and a good case can be made for its use" (p. 10). Other systematic reviews and meta-analysis, such as those by NICE, among others, have restricted evidence to exclude evidence from small *N* experiments. Thus, NICE's (2012) clinical guideline on adults with autism excluded all small *N* experiments with adults with autism, leading to conclusions that differ from the conclusions in other meta-analyses and systematic reviews of small *N* experiments of interventions with adults with ASD (Bishop-Fitzpatrick, Minshew, & Eack, 2013; Chowdhury & Benson, 2010; Palmen, Didden, & Lang, 2010). For example, whereas NICE (2012) did not discuss the use of behavioral interventions, Palmen et al.'s (2010) systematic review identified 20 small *N* experiments with 116 participants which evaluated social, academic, vocational, and domestic skills. They concluded that "behavioral interventions can be successfully used to improve adaptive skills in young adults with high functioning ASD" (pp. 614–615).

Similar differences in definitions of EBPs can also be seen in definitions of EBPs using small *N* experiments. For example, when defining EBP in special education, Horner et al. (2005) proposed that "A practice may be considered evidence based when (a) a minimum of five single-subject studies that meet minimally acceptable methodological criteria and document experimental control have been published in peer-reviewed journals, (b) the studies are conducted by at least three different researchers across at least three different geographical locations, and (c) the five or more studies include a total of at least 20 participants" (p. 176). These criteria are more stringent than Chambless and Hollon's, and, thus, a treatment might meet criteria as an EBP using one set of criteria, but not when using another set.

Some definitions of EBP distinguish multiple levels of evidence. For example, Rusch and Frances' (2000) *Expert consensus guidelines for treatment of psychiatric and behavioral problems in mental retardation* distinguished three levels of evidence. First-line treatments were "options that the panel feels are usually appropriate as initial treatment for a given situation" (pp. 162–163). Second-line treatments "are reasonable choices for clients/patients who cannot tolerate or do not respond to the first line choices" (p. 163), and third-line treatments "are usually inappropriate or used only when preferred treatments have not been effective" (p. 163). They defined these levels of treatment operationally as scores on a 9-point Likert scale as rated by an expert panel where first line was defined as at least 50% of the panel rating the treatment as a 9 and second-line treatment's entire confidence limits had to be greater than 6.5. A third-line treatment was one in which a portion of the confidence limit fell below 3.5.

Some approaches to EBP propose treatment algorithms and stepped-care approaches which recommend different treatment intensities—often with increasing costs—as one proceeds from at-risk populations (primary or secondary prevention) to severe, chronic treatment-nonresponsive problems (clinical treatment). For example, Rusch and Frances' (2000) guidelines identified first-, second-, and third-line treatments, suggesting that practitioners should use these guidelines to sequence treatment options. In the literature on offenders, treatments are sometimes developed according to responsivity–need–risk model in which clients with higher risk for offending receive more intensive treatments.

Rationale

There are a number of reasons why we should be interested in EBP. These can be grouped under two categories: professional and personal ethics (Schreck & Miller, 2010) and economics and efficiency.

Professional and personal ethics

Standards of professional and personal ethics require practitioners to adhere to the principles of beneficence and nonmalfeasance, namely, doing good and avoiding harm to the client. In addition, therapists should be competent to treat their clients effectively, select treatments based on scientific evidence (where available), appraise the likely effects of all alternate therapies, have specific treatment objectives, and cooperate to reduce conflicts with other professionals (Schreck & Miller, 2010). Finally, clinicians should be honest with their clients, which includes giving their clients an honest appraisal of the likely response to treatment, benefits, and risks of harm, including unknown risks of harm for unevaluated therapies (APA, 2010). Service funders too have ethical obligations to maximize client benefit and the funders by investing resources in treatments that are effective and in an equitable fashion. Thus, professional ethics for practitioners and administrators points in the direction of implementing EBPs to increase client gains, avoid client harms, and use limited resources wisely.

Psychotherapy economics and personal ethics intersect. Individual therapists and services implicitly and explicitly make choices over limited resources every moment at work. (These choices may be influenced by other factors such as funding and number of treatment sessions allocated by supervisors or state-approved plans.) They choose how much time to spend with each client, when to take on new clients, and when to leave others on waiting lists or when to seek out new clients. They allocate clients to high-cost individualized treatment by professionals or choose to allocate them to cheaper therapies delivered in groups by other therapists, perhaps with more limited therapeutic skills. They choose between investing time in continuing education, thereby reducing client treatment, with the possible benefit to clients in the future if the therapists become more competent in treating important, common, treatment-responsive problems. They engage in administrative, research, and other activities and with no immediate client outcome, but with the possibility of distant client benefits. None of us can escape the economic implications for psychotherapy practice!

Psychological treatments sometimes harm clients

It is tempting to believe that psychological treatments do not harm clients, but this is untrue. Lilienfield (2007) lists several harmful psychological treatments including one related to developmental disabilities. These included facilitated communication, which has the potential to result in false accusations of child abuse against family members.

A second recent comes from the use of psychotherapy and autism in France. The BBC (Schofield, 2012, http://www.bbc.co.uk/news/magazine-17583123) reported that in France psychoanalysis and psychiatry dominates the treatment of autism. This model maintains Bettelheim's psychoanalytic model that autism is caused by so-called refrigerator mothers (Bettelheim, 1967). Consequently, the treatments offered include wrapping the child in cold sheets to reconnect them with their partially clothed bodies ("packing"). This may result in children crying and struggling and has been implemented without parental knowledge for extended periods of time in some cases (Cheng, 2012). Other psychoanalytic treatments for autism include psychoanalysis for the family. In France, only approximately 20% of children with autism receive education, and of these, many receive only part-time education. Schofield suggested that psychoanalysis results in denial of education and effective treatment, very high rates of family divorce, and single mothers left alone to manage their child with autism without help. Currently, there is no evidence that any of these psychoanalytic treatments are effective, although psychoanalysts offer uncorroborated anecdotal reports of "spectacular" results. According to Cheng, Marie Dominique Amy, president of French psychotherapy and psychiatry association, claimed that after packing, "Amy said she had seen autistic children start talking, writing and drawing after the sessions." There are no empirical studies supporting these claims.

Failure to provide effective treatment

There is a much more insidious and ubiquitous form of harm: failure to provide effective treatment. We often observe high-need clients who receive no treatment, who receive unnecessarily extensive assessments, or who are referred to other services that do not materialize for months or sometimes a year. A more subtle problem is that services do indeed respond to individuals with high needs but do so by preventing the opportunity to engage in challenging behavior. For example, a high-need individual might be placed on one-on-one staffing, separated from a group, or from materials or other people so they cannot engage in challenging behavior. This strategy may be legitimate as a first, short-term strategy to prevent harm, but, as a long-term strategy, it denies effective treatment to the person resulting in both a restricted lifestyle and harm to the individual and others because the challenging behavior goes untreated.

After a choking incident, for example, an individual with life-threatening pica might be placed on one-on-one staffing, and all objects that they might ingest might be removed. Such strategies avoid client harm in the short term, but this individual might benefit from a number of effective, nonaversive EBPs such as noncontingent reinforcement and environmental enrichment based on functional assessment and analyses (McAdam, Beidbord, Dahl, & Williams, 2012) (see meta-analyses of pica treatment by Hagopian, Rooker, & Rolider, 2001; McAdam et al., 2012; McAdam, Sherman, Sahaddon, & Napolitano, 2004). If, however, the individual does not receive effective, evidence-based interventions, then the individual will continue to live in a highly restrictive environment indefinitely and will not receive effective treatment to increase appropriate behavior and decrease pica. Thus, in such circumstances, this individual may be kept safe for many months until they are then engage in life-threatening pica again.

The *Right to Effective Treatment* statement (Van Houten et al., 1988) addresses the issue of access to effective treatment. Some aspects of this statement—such as access to a responsive environment and services whose overriding goal is personal welfare—are noncontroversial, even if many services struggle to do these things. The right to the most effective treatment procedures available, however, is most controversial because (a) the statement labeled some nonaversive interventions as unacceptable due to the lack of treatment effect or treatment that produces change too slowly and (b) it opened the door to "quicker acting, but temporary more restrictive procedures" (p. 383) (see Table 1.2 for a summary of Van Houten et al.). This issue continues to be recognized in contemporary statements on restraint and restraint reduction in which calls for judgments concerning the risk to benefit ratio, including the risks of ineffective treatment and the harm and distress that accrues the person from not treating and treating ineffectively and the possibility that a more intensive intervention that reduces harm and results in better client outcomes (Applied Behavior Analysis International, 2010).

Some saw the *Right to Effective Treatment* statement (Van Houten et al., 1988) as an open door for the use of aversive treatments, such as contingent electric shock, an issue that still resonates today. Thus, the *Right to Effective Treatment* statement and other events in the 1980s stimulated the growth of organization such as TASH and the Positive Behavior Support movement to promote effective, nonaversive treatments.

Economics and efficiency

The cost of services has always been an important consideration in determining services available to people with developmental disabilities. Economic analysis of mental health services generally (Knapp & McDaid, 2012) assumes that mental health systems should alleviate symptoms, promote quality of life, support caregivers, and improve life chances but notes that this is done within the context of

Table 1.2 A Summary of the Right to Effective Treatment Statement (Van Houten et al., 1988)

Right	*Examples*
1. Therapeutic environment	Physical and social environment is safe, humane, and responsive and has an acceptable standard of living. The individual has access to therapeutic and leisure services and leisure and enjoyable materials. Activities and materials should respect client choices and be age appropriate. There should be frequent positive interactions to ensure enjoyment, learning, and independence with the fewest restrictions necessary to ensure individual safety and development
2. Personal welfare	Behavioral treatment to promote functional skills and independence, immediate, and long-term welfare with active individual or proxy participation. Risk and professional performance is overseen by human rights and peer review committees to impose community and professional standards
3. Treatment by a competent behavior analyst	Person delivering services must have academic training and clinical competence. For complex problems, a doctoral-level behavior analyst is needed to ensure appropriate assessment, treatment, training to service providers, consultation, and follow-up
4. Programs that teach functional skills	Services have the goal to increasing individuals' effective functioning to teach adaptive behavior, behavior to terminate aversive stimulation, reduction in behavior that is dangerous and restricts access to independence and social acceptability, and behavior that is beneficial to society. All individuals should be treated as capable of learning. This may include temporary exposure to discomfort and risk
5. Behavioral assessment and ongoing evaluation	A diagnostic assessment and a behavioral assessment that includes interviews, direct observation, and incorporation of behavior assessment information into a treatment plan. Ongoing evaluation using objective, public data
6. Most effective treatment available	Effective and scientifically validated treatment. Individual and public educated concerning most effective treatment. Restrictive procedures are unacceptable unless shown to be safe and effective. Nonrestrictive interventions are unacceptable when more effective, faster-acting alternatives are available

limited resources that must be allocated carefully to how the resources are used to achieve client outcomes. For many products, such as groceries, consumers are routinely exposed to purchasing the product and can readily judge the cost and quality of the product. Mental health services are not like groceries because people purchase them infrequently, it is difficult to judge their quality, and they often come with perceived stigma. Additionally, often the consumer themselves do not directly purchase them; rather, a third party such as a family member, guardian, or service provider purchases those products on behalf of the consumers. In some cases, such as offenders, courts impose services and treatment on clients. Resources for mental health come from prepayment systems, such as taxes, salary-based contributions, voluntary health insurance schemes (also known as private health insurance), and out-of-pocket expenses used to purchase services on the basis of efficiency, that is, achieving the maximum relevant outcome for the price, and equity, that is, fairness of the distribution of outcomes, access, and payment across individuals and society.

Cost-effectiveness analysis is one approach to economic evaluation of human services which asks two questions: "Does it work?" but also "Is it worth it?" Cost data include staff salaries in prevention, treatment, and other services; facility costs, such as cleaning; overheads, such as personal; and capital costs, such as buildings. Effectiveness can be measured by using common metrics, such as changes in standard measures of depression. Effectiveness can also be measured using quality-adjusted life years (QALYs), a measure of years of perfect quality of life and the cost of increasing the number of QALYs due to treatment. Some cost-effectiveness evaluations use monetary values by comparing the total costs and benefits of two or more treatment options, including no or minimal treatment.

Decisions to implement EBPs are made both individual and societal levels. For example, a practitioner might decide to increase the number of sessions given to one client and reduce the number of sessions to another. Alternatively, an agency, government, or insurance provider might decide to stop funding early intervention and provide sensory integration therapy within a certain agency. (Chapter 4 discusses this issue in more detail.)

Methods Used in Evidence-Based Practice

Evidence-based approaches are characterized by systematic reviews and meta-analyses. These approaches use (a) explicit statement of research questions; (b) efforts to make procedures such as literature searches transparent by describing search strategies, inclusion, and exclusion criteria; (c) in the case of meta-analysis, objective methods of measuring, combining, and disaggregating treatment ESs; (d) clear rules for describing the strength and quality of evidence; and (e) dissemination and evaluation of EBPs to practice through professional training and support and service design or application to individual cases. The next sections review each of these five aspects of EBP.

Formulating a research or clinical question

Why does someone conduct a systematic review or meta-analysis? Why does a practitioner ask a question concerning the application of EBP to a particular case? What question are we trying to answer?

Schlosser, Wendt, and Sigafoos (2007) noted that the purpose of systematic reviews and meta-analyses is sometimes unstated or unclear. For example, if a paper states that its purpose is to "summarize" or "review" the literature on a particular treatment or problem, it can readily become unfocussed and the research question does not guide the authors how to search the literature. To solve this problem, Schlosser et al. recommended stating research questions using this three-part format: subject (the participants), verb (the intervention), and object (an outcome). Thus, a better question than the original research question would be: "In adolescents with mild ID, what are the effects of job coaching, on chances of employment in integrated settings, income, and comprehensive treatment costs?"

Clinicians also face the challenge of formulating clear, answerable questions when attempting to identify the best treatment for a specific client using EBP (Straus et al., 2011), but how should a clinician formulate such as question? For example, consider a clinician working with a 50-year-old woman with mild ID and a diagnosis of borderline personality disorder who presents challenges to the staff because she refuses to comply with preventative medical treatment for poor circulation in her legs. If the problem goes untreated, there is a possibility that she may have infections, ulceration, or even amputation, but when the staff ask her to engage in any rehabilitation, she curses at them, scratches her legs, destroys the equipment, and throws things at the staff. What question should the clinician ask? They might ask "well what do I do?" which might be a general cry for help from someone in a difficult situation who does not know the best course of action. They could ask questions about the best treatment for noncompliance in adults with mild ID, or they could ask what the best treatment for borderline personality disorder is. (In the end, since the clinician could find no literature specifically on the treatment of borderline personality disorder in adults with mild ID, the clinician simplified the question and searched for EBPs for borderline personality disorder. She went on to modify those procedures for someone with mild ID.)

Systematic searches

To address the issues of bias or incompleteness of literature included in traditional narrative reviews, systematic reviews and meta-analyses attempt to describe their search strategies completely and with transparency. Search strategies may include online searches; searching of reference sections of articles retained from the searches and/or from review articles, book chapters, etc.; searching of citations of articles; online searching of journal contents; contacting authors and/or leading researchers; and hand searches of journals. Online searches may vary along a number of dimensions including (a) the number of databases searched; (b) the variety of databases search, for example, whether the databases include primarily psychological, educational, special education, medical, rehabilitation, theses, or other content areas; (c) the range of years searched; (d) the choice of disability-related and other search terms and whether supplementary searches are conducted after the initial search; (e) the inclusion/exclusion criteria; and (f) whether or not and how raters are trained to search the resulting abstracts and the reliability and validity of their searching. Most emphasis has been placed on the reliability and transparency or reporting searches, but validity has received less attention.

For example, some systematic reviews include flow diagrams showing the numbers of papers located and retained various stages of the searches, and some report the reliability of doing so by having more than one person to conduct the search. Some searches also report an explicit procedure to resolve whether to include papers where raters disagree on inclusion.

Searches also vary in terms of their difficulty. For example, searching for a very specific intervention in a very specific population, such as the effects of Active Support on staff and client behavior (Hamelin & Sturmey, 2013), is relatively easy because there is a unique string to use ("Active Support") and a small literature with relatively few authors. Thus, it is relatively easy to conduct a complete search. On the other hand, broad searches, such as psychosocial treatments for internalizing behavior disorders, would be quite difficult to do because the literature is large and dispersed across many sources, and there are many synonyms for both "psychosocial treatments" and "internalizing disorders." Thus, such a search would probably effortful and expensive and also impossible to conduct a complete search.

The issue of validity is illustrated by Duggan, Morris, and Adams (1997) who evaluated the accuracy and completeness of online search strategies. First, they hand searched the *Journal of Intellectual Disability Research* from 1957 to 1994 and identified 56 RCTs in that journal. None used the word "randomized" in the title, and a mere nine used it in the abstract. Of 37 RCTs published between 1974 and 1994, 3 and 37 were in Psychlit and MEDLINE, respectively, although MEDLINE contained one incorrect abstract. When they evaluated online searches of the journal to identify RCTs, all searches identified only a few of the RCTs, perhaps reflecting problems with poor-quality indexing and keywords. Duggan et al. show that online searches alone may result in incomplete samples of the literature and researchers should supplement them with other search methods, such as hand searchers of references sections and journal content pages.

Meta-analysis

One characteristic EBP methodology is meta-analysis, a family of statistical procedures that calculate numerical measures of ESs and combine data from multiple studies. Meta-analyses use these data to estimate the average ES for relevant questions, such as the average ES for a particular clinical problem and treatment, or more focused questions, such as the average ES for specific treatments for specific problems. As a field develops, important questions may change from "Does any treatment work for this problem?" to "Does this specific treatment work for this problem?" to "Does this specific treatment work for this problem in this specific population?" to "What is the relative effectiveness of one treatment over another?" to "How does this new treatment compare to existing, standard treatment treatments?" Which question is relevant depends on the extent of the literature in a particular field. For example, within the field of treatment of depression, there are over 200 RCTs permitting answers of some focused specific questions (Cuijpers et al., 2012). For borderline personality disorder, there are few RCTs that have evaluated only two treatments (Sneed, Ferteck, Kanellopoulos, & Culag-Reinlieb, 2012) permitting answers to only general questions as to treatment effectiveness and the relative effectiveness of only two psychosocial treatments. Finally, for somatization disorder, a recent review identified no adequate RCTs or small *N* experiments; thus, there was no experimental evidence available to guide practice (Allen & Woolfolk, 2012).

Meta-analysis methods have been developed for both RCTs and small *N* experiments. There are a wide range of ES measures. For RCTs, the most commonly used measure is Cohen's *d*, which expresses the differences between the means of experimental and control groups as a *z*-score, usually using the standard deviations of the pretreatment scores in both groups. In meta-analyses of small *N* experiments, the most commonly used metric is the proportion of nonoverlapping data points between treatment and baseline, although many other ES measures exist for small *N* experiments.

Meta-analysis was originally applied to experimental psychology in the late 1950s (Glass, 2000). In response to critiques of the effectiveness of psychotherapy (Eysenck, 1957) and the limitations of unsystematic narrative review, meta-analysis was rapidly adopted as a tool to evaluate psychotherapy outcome research (Smith & Glass, 1977). Meta-analysis is now widely in many fields. Many professional bodies and government agencies now use systematic reviews and meta-analysis to ensure that clients have access to the most effective and efficient therapies and are protected from harmful treatments. Consequently, they have appointed expert panels and even funded government agencies, such as NICE in the United Kingdom, to use EBP methods to direct professional practice and government funding. Such initiatives produce the lists of therapies that are deemed to meet criteria for EBP such as those from NICE, Cochrane reviews, and the APAs and pronouncements from professional bodies, as to therapies that are evaluated and shown to be effective, ineffective, or harmful and unevaluated therapies. Such pronouncements often grade therapies in terms of the strength of the evidence to support conclusions.

The results of meta-analyses, especially those from government and professional bodies, are often translated into practice guidelines. These guidelines indicate which treatments should be preferred and which should be avoided and in which order practitioners should implement treatments. Often, practice guidelines, sometimes referred to as stepped-care approaches, begin by recommending simple, cheap, preventative effective treatments, if they are available. Next, they may move onto more expensive but more effective treatments, perhaps delivered in groups by nonclinical staff or nonspecialist therapists. The last stage of guidelines may be specialist therapies that are costly, delivered by highly trained individual therapists for relatively treatment-resistant problems. Sometimes, practice guidelines may include relatively expensive treatments that have limited effectiveness because other better alternatives have already been attempted (see NICE, 2012, for examples of practice guidelines for adults with ASD based on systematic reviews and meta-analyses).

Quality of evidence

A most vexing question is that of quality of evidence. If the study quality is low that may reduce our confidence that change occurred and/or that the treatment caused the change and thus mean that we cannot conclude that treatment is an EBP. EBP criteria give preeminence to experiments because they enhance the probability that observed changes were due to treatment and not to other factors such as to the effects of the treatment and not to chance, passage of time, the possibility that assessment alone resulted in change, or inadvertent confounding of different types of clients with treatment conditions (Campbell, 1969; Kazdin, 2010). Nonexperimental studies, such as narrative case studies, AB single case reports, pre–post group designs, and group designs without randomization, are usually seen as poor-quality studies. A series of nonexperiments—no matter how many or difficult it may be to conduct experiments—can never exclude factors other than treatment as explanations of change.

Quality of randomized controlled trials

For many, RCTs are the “gold standard” of evidence, but not all RCTs are well designed, and we cannot always be confident in the conclusion that the treatment caused the change in some important outcome. The list of requirements for a well-designed RCT is long, making it expensive and effortful to conduct a well-conducted RCT; perhaps no study can meet all requirements.

CONSORT standards One of the most recent sets of codified standards is the 2012 CONSORT standards (Moher, Schulz, Altman, & the CONSORT Group, 2010). One of the important aims of CONSORT standards is transparency. The authors should report all relevant information including a flow diagram for all participants who entered the RCT including the numbers (a) assessed for eligibility, (b) excluded from treatment with reasons for exclusion, (c) randomized to each intervention, (d) who did not receive the assigned intervention and the reasons, (e) lost to follow-up and the reasons the number of participants who discontinued intervention and the reasons, and (f) included and excluded in analysis and the reasons. When an RCT reports all these numbers, it is to evaluate the quality of the RCT and judge the effects of recruitment and inclusion on external validity, the effects of attrition, and statistical analyses on internal validity.

The 2010 CONSORT standards generated 25 standards for reporting an RCT and set of standards for the abstract of an RCT. Moher et al. (2010) also justified and illustrated each standard. For example, there are four standards (numbers 8a, 8b, 9, and 10) related to randomization. Standard 8b described that the report should describe the type of randomization and details of any restriction used during randomization so that it is unclear if the authors truly randomized participants. For example, if an author states that they randomized participants to two groups, it is unclear how they did that; however, if they state that they block randomized consecutive groups of three participants using a constrained randomization sequence generated by Stat 9.0 (StataCorp, College Station, TX), then it is clear what they did and the threats to internal validity are lessened. The reader is referred to Moher et al. for a complete listing of CONSORT standards.

Grading of recommendations assessment, development, and evaluation Guyatt et al. (2010) aim to provide “a highly structured, transparent, and informative system for rating quality of evidence” (p. 996). GRADE rates the quality of evidence of studies that are included in a meta-analysis by rating the quality of the RCT, consistency of findings across RCTs, if the available evidence is directly related to the question at hand, the degree of precision that the data allow, and any publication bias. An important consideration in GRADE is rating of the significance of the dependent variable, since some dependent variables, such as mortality, are more important than other outcome variables, such as flatulence. The degree of precision refers to the magnitude of treatment effect and the sample sizes. Thus, if treatment effects are very large in several RCTs with large samples, we can be confident that the treatment is robustly effective, and if the dependent variable is important, then we can attach greater social significance to the findings.

Researchers rarely considered the social significance of the dependent variable in meta-analyses in developmental disabilities, but this may be very important. For example, in considering the effectiveness of interventions for anger and aggression, we might place different degrees of significance to different outcomes, such as acquisition of adaptive skill, aggression, injuries to others, and quality of life. An example of the application of GRADE to treatment and developmental disabilities is NICE’s (2012) guidelines for the treatment of adults with autism.

Quality of small N *experiments*

As with RCTs, small *N* experiments vary widely in their quality and their ability to let us confidently conclude that the treatment caused the change in some socially important behavior. Horner et al. (2005) defined the quality of small *N* experiments as follows: “Single-subject research documents a practice as evidence-based when (a) the practice is operationally defined; (b) the context in which the practice is to be used is defined; (c) the practice is implemented with fidelity; (d) results from single-subject research document the practice to be functionally related to change in dependent measures; and (e) the experimental effects are replicated across a sufficient number of studies, researchers and participants to allow confidence in the findings.” They then went on to define each of these five criteria in further detail.

The National Center on Autism’s *National Standards Project* on EBP and autism also developed a quality checklist, the *Scientific Merit Rating Scale* (SMRS), for the quality of small *N* experiments. The SMRS has five domains: “(a) research design,

(b) measurement of the dependent variable, (c) measurement of the independent variable or procedural integrity, (d) participant ascertainment, and (e) generalization" (p. 16). The SMRS generates a 6-point rating from 0 to 5 of study quality with operational definitions of each level of evidence for each of the five domains listed earlier. Consider a study with an SMRS rating of 3. In terms of research design, such a study would have at least two comparisons of control and treatment conditions, at least 3 data points per condition, and at least two participants, and some data loss would be possible. In terms of measurement of the dependent variable, there might be no calibration data to ascertain the absolute accuracy of the data, interobserver agreement (IOA) might be better than 80% or a kappa greater than .4, IOA might be collected for at least 20% of the data, and data might only be collected in the treatment condition. With respect to measurement of the independent variable, treatment accuracy might be at least 80%, implementation data might be taken in 20% of parts of sessions, and no treatment fidelity IOA might be reported. In terms of participant ascertainment, diagnoses might be confirmed by an independent professional, or the study might use blind evaluation using at least one psychometric instrument, or an independent qualified diagnostician might use DSM criteria. Finally, treatment generalization might either use objective data with some maintenance or collect generalization data across at least some settings, stimuli, or persons. The National Standards Project (National Autism Center, 2009) developed a coding manual to assist in the coding of the quality of the studies, and they trained reviewers to a criterion of at least 80% IOA. After initial training on one small *N* article, they also collected IOA on coding on one article by each reviewer and coding IOA remained above 80%. The National Autism Center (2009) developed a *strength of evidence* classification system which judged the strength of evidence across available studies with four level of strength of evidence including a final category to distinguish ineffective from harmful treatments (see Table 1.3).

Strategies for handling study quality

Studies that may enter into systematic reviews and meta-analyses vary in their quality. To address this, systematic reviews and meta-analyses have commonly used two strategies for handling study quality. The first is to apply relatively stringent inclusion criteria during the literature search phase. For example, a systematic review of group designs might only include RCTs and exclude pre–post designs. Similarly, a systematic review of small *N* experiments might exclude AB designs since they are not experiments or require a certain number of baseline and intervention data points. Although it is possible to set the bar higher—for example, by insisting that only papers with treatment integrity or follow-up data are included—few systematic reviews and meta-analyses have done so, probably because in many cases there would be little literature left to review.

The second strategy is to include studies of varying quality and report ESs broken down by study quality. For example, a meta-analysis might report ES for experiments and nonexperiments or for "low"- and "high"-quality studies as measured on a quality checklist. If both low- and high-quality studies agree on the ES, one may be

Table 1.3 A Summary of the Four Levels of Evidence from the National Standards Project

Level of evidence	*Definition*
Established	1. Two group designs or four small *N* experiments with at least 12 participants with no conflicting results or at least three group experiments or six small *N* experiments with a minimum of 18 participants with no more than one study reporting conflicting results 2. Had SMRS scores of 3 or greater 3. Reported beneficial treatment effects for specific targets 4. These results may be supplemented by other lower-quality studies
Emerging	1. One group design or two small *N* experiments with a minimum of six participants with no conflicting results 2. SMRS scores of 2 3. Beneficial treatment effects on one dependent variable for a specific target 4. These studies may be supplemented by those with higher or lower SMRS scores
Unestablished	1. May or may not be based on research 2. Had beneficial effects reported on very poorly controlled studies with SMRS scores of 0 or 1 3. Have claims based on testimonials, opinions, or speculation 4. Were ineffective, unknown, or adverse treatment effects based on poorly controlled studies
Ineffective/harmful	1. Had two group designs or four small *N* experiments with at least 12 participants with no conflicting results or at least three group experiments or six small *N* experiments with a minimum of 18 participants with no more than one study reporting conflicting results 2. SMRS scores of at least 3 3. No beneficial treatment effects for one dependent measure for a specific target behavior or had adverse treatment effect reports on dependent variable

Adapted from National Autism Center. (2009). *National standards project*. Randolph, MA: Author.

more confident on the magnitude of the ES than otherwise; however, if low-quality studies produce larger ESs than poor-quality studies, one might be more conservative and only use the estimates of ES from good-quality studies.

Some systematic reviews and meta-analyses use the scores on quality checklists to split studies into “high”- and “low”-quality studies. This may not be a good idea, as some checklists give an equal or greater number of points to how an abstract is written as to whether or not the RCT randomized adequately. The former is nice, but does not threaten the ability to conclude if the treatment caused the change, but the latter is an essential feature of an RCT that is probably fatal to the ability to answer this question. Some systematic reviews and meta-analyses have addressed this problem by defining "high"-quality studies as those that have all of several features (e.g., reporting ES separately for small *N* experiments with reliability, experimental control, treatment integrity, generalization, and social validity data).

Small *N* experiments: in or out?

There is a considerable disagreement over the relative status of RCTs and small *N* experiments. In fact, many authors refer to RCTs as the "gold standard" of evidence. This is unsurprising as the conventions of group designs are the foundation of evaluation of drugs and, by extension, other medical procedures. Psychology's adoption of the group design, hypothesis testing, and statistical testing as the foundation of knowledge in experimental psychology places high value on RCTs. The convention in hypothetico-deductive psychological science is that hypotheses flow from extant theories, which are tested in experiments and theories are confirmed or modified, dependent upon the results of experiments (Chiesa, 1994). Indeed, some have referred to RCT position in behavioral science as hegemonistic (Keenan & Dillenberger, 2011).

Some authors and organizations such as the NICE and Cochrane reviews explicitly exclude small *N* experiments from consideration. Thus, several reviews of EBP and developmental disabilities have concluded that early intensive behavioral intervention is not an EBP (e.g., Maginnis, 2008). Others downgrade evidence from small *N* experiments by describing them incorrectly as "case studies" or "prescientific studies." Others have suggested that the external validity from small *N* experiments is problematic due to the small number of participants, which is true unless researchers conduct direct and systematic replications, which has in fact occurred in many areas.

Others have criticized RCTs and defended small *N* experiments equally vociferously. Chiesa (1994) has pointed out the basic flaws in the logic of RCTs in which experimenters conduct a group design to infer causality between the independent and dependent variables and make this inference based on statistically significant changes in group mean scores. Inspection of individual scores, however, may reveal that only some—perhaps only a few—participants did indeed change; many may remain unchanged; some may get worse and some get worse because of the treatment. Given that many participants do not change and some changed for the worse, it is illogical to conclude that the treatment caused the improvement, since for some portion of the participants, the improvement did not occur.

Most experimental psychologists and those involved in psychological therapy outcome research ignore these conceptual objections, but there are also procedural problems. Namely, part of the rationale for RCTs is that the experimenter defines a population, the experiment is conducted on a random sample from that population, and the results from that sample are then generalized to the population. Of course, almost all experiments do not define the population and for that and for other practical reasons cannot randomly sample from the undefined population. Thus, even if we ignore Chiesa's logical objections to RCTs, we cannot ignore the impracticality of conducting RCTs in the manner in which they are supposed to be conducted and the limits to generalization from RCTs which are not based on random samples.

Another important question for EBP is to predict the best treatment for specific individuals (Chiesa, 1994; Keenan & Dillenberger, 2011). This indeed is the question that clients and therapists ask: They do not ask if the average person will benefit from this treatment. They ask if *this specific person here and now* will benefit from this treatment for this problem. Since the mean score of the treatment group in an RCT predicts individual response to treatment so poorly, RCTs offer limited guidance on treatment selection for individual clients.

Other issues

Grouping treatments

When conducting systematic reviews and meta-analyses, how should the authors label treatments? Is it a video modeling and behavioral antecedent procedure or a social learning intervention?

One approach to grouping therapies is to do so empirically. Smith and Glass (1977) used multidimensional scaling to group therapies into four "superclasses": behavioral, psychoanalytic, Gestalt–Rogerian, and rational emotional therapy/transactional analysis. To date, there are no such examples related to developmental disabilities, but such data might be interesting. Such "superclasses" of therapies are crude and omit distinctions that are very important to some researchers and practitioners. For example, some might make a very clear distinction between behavior analysis, behavior modification, PBS, and cognitive behavioral therapy (CBT); others might readily lump them together. Reporting agreement of coding of papers into therapy types addresses only the reliability but not the validity of coding.

One good example of this is the status of client self-talk. Suppose client learns to say "Stop! Walk away and relax" (loud or privately) when someone irritates them and, once they have done so and they do indeed relax, say to themselves "That was good. I am relaxed now. That was the adult thing to do." What kind of therapy is this? It is tempting to call it cognitive therapy since it involved verbal self-instruction and self-praise. Behavior analysis, however, also contains an account of self-control in which a person learns to modify their own behavior by emitting one behavior, the controlling response, to influence the future probability of another response, the controlled behavior. This behavioral model places emphasis on the variables that influence the emission of the controlled response, such as its establishing operations, discriminative stimuli, and consequences (Skinner, 1953). Cognitive therapists view such procedures as evidence of covert private behavior and changes in alleged cognitive structures, whereas behaviorists see such procedures as examples of self-control (Skiner, 1953; Sturmey, 2006a, 2006b). Thus, both parties might agree if the procedure is effective but would disagree on the status given to the observed behavior, the model of causality for the observed change in behavior, and the type of therapy.

A final aspect of this problem is that when there are combinations of procedures, such as CBT, different authors place different emphasis or infer efficacy for only one component of the package (Ward-Horner & Sturmey, 2012). For example, CBT for depression is an EBP for depression which includes behavioral activation and cognitive restructuring, but behavioral activation alone produces similar ESs to the CBT package (Cuijpers et al., 2012; Sturmey, 2009a), suggesting that behavioral activation alone is the effective component of CBT for depression. Similar concerns have been raised over the debate over the effectiveness of CBT with people with developmental disabilities in areas such as CBT and anger management, where it is unclear what the contribution is from nonbehavioral components (Sturmey, 2006a, 2006b, 2006c; Travis & Sturmey, 2013), although there is not agreement among professionals on this point.

Flatulence or mortality?

If two treatments have the same ES, are they equally valuable? Perhaps not. Treatments that have a modest reduction in mortality are probably more socially and personally significant than those that reduce flatulence. Thus, a significant but neglected problem is evaluating the social or clinical significance of dependent variable(s).

Meta-analyses emphasize the magnitude of ESs, not the importance of the outcomes. This problem is apparent in the literature on developmental disabilities. For example, Chapter 2 reviews the meta-analyses of skills training interventions, some of which relate broad impact on the rate of development and the possibility of returning children with ASD to typical functioning and not consuming special education, mental health, or residential services. Other meta-analyses relate to more narrow dependent variables, such as learning to shop or use technology. Such outcomes may be personally significant but may often be personally and financially less valuable than interventions that have broad impact on development and which may lead to removal of diagnoses such as autism. This problem is also evident in Chapter 12 of this volume on offenders, which only found data from experiments on short-term immediate behavioral outcomes, but no did not find data from experiments on reoffending, although such data are available from nonexperiments. Thus, the evaluation of individual systematic reviews and meta-analyses requires consideration of the value of the dependent variables used in the experiments that go into the meta-analyses.

Application and dissemination

Recall that the aim of EBP is for practitioners to apply the best treatment to actual clients effectively to achieve socially significant outcomes for each client. This can be done at the level of individual cases or entire services. There are few examples of the former with individuals with developmental disabilities, although Schlosser and Raghavendra (2004) illustrated how to formulate a clinical question for an individual case and use systematic reviews and meta-analyses to answer the questions about treatment options (see Chapter 2). Additionally, Straus et al. (2011) developed a useful manual for the application of evidence-based approaches to medical problems, which can be used successfully to identify effective EBPs to use with individuals with ID or autism.

There are several examples of disseminating EBPs. Perhaps the best known is Britain's Improving Access to Psychological Therapies Project to disseminate EBPs for anxiety and depression in the British National Health Services. Following NICE systematic reviews and meta-analyses that identified CBT as an EBP for depression and anxiety, NICE developed treatment guidelines, and an economic analysis indicated the possibility that CBT would result in significant cost savings (Layard, Clark, Knapp, & Mayraz, 2006). Thus, NICE identified a common, treatment-responsive problem which could be treated effectively with manualized treatment by many commonly available and trainable therapists that would result in large-scale personal benefits and economic benefits for society. Subsequent demonstration projects showed that large-scale implementation is possible (Clark, 2011) but required considerable effort to train and maintain therapist consistency in implementing CBT and systematic monitoring of client outcomes. Foa, Gillihan, and Bryant (2013) describe multiple projects to disseminate prolonged exposure therapy for PTSD in multiple locations across the world and the adaptations they made to numerous local contexts.

Perhaps the closest to such large-scale dissemination in developmental disabilities is Willner et al.'s (2013) recent evaluation of cognitive behavioral anger management. The trial was a multisite trial which involved multiple local therapists and 179 individuals with ID and anger management problems randomly assigned to CBT or treatment as usual. Based primarily on staff ratings, the treatment group had less anger-related problems and more use of coping skills following treatment. Interestingly, this trial did measure treatment integrity and individual attendance at therapy, both of which were highly variable and sometimes low. It indicates that it is possible to implement CBT for individuals with ID using local care staff and a treatment manual over multiple sites but also suggests that such an approach needs to be improved with greater attention to therapist training, treatment integrity, client attendance, and perhaps other interventions for nonresponders.

Evidence-Based Practice and Developmental Disabilities

EBP has until recently been relatively neglected. This has led some to opine that there is no evidence but only compassion available to guide practice (King, 2005). Others have suggested that there is insufficient evidence to guide treatment of mental health problems in adults with disabilities (Hastings, Hatton, Lindsay, & Taylor, 2013) or that there is insufficient evidence to prefer one therapy over another (Emerson, 2006)—the so-called Dodo Bird hypothesis. These claims contrast with numerous systematic reviews and meta-analyses (see Chapters 2 and 3) identifying

EBPs and some economic analyses demonstrating the economic benefits of some interventions with people with ID/autism (see Chapter 4).

Why is there controversy? First, many common practices have little or no support, including sensory integration therapy, TEACCH, social stories, psychotherapy, and counseling. Practitioners and advocates for these treatments and people who genuinely believe that such treatments are effective are naturally put on the defensive by claims of lack of evidence. Second, there is a disagreement over the status of RCTs and small *N* experiments. As this book demonstrates, the majority of evidence comes from small *N* experiments. Sometimes, small *N* experiments are described incorrectly as “prescientific” or “case studies.” Sometimes, they are criticized because of the small number of participants and hence the apparent lack of generality, but such criticism fails to take note of the role of direct and systematic replication in making generalization (Sidman, 1960). Sometimes, they are criticized because of the small number of studies that report generalization and maintenance data. While it is true that only about a third of small *N* experiments do so, almost no group designs evaluating the nonbehavioral treatments report any data on generalization, treatment integrity, or social validity. So, for those who do not accept small *N* experiments as experiments, there is indeed not much evidence left when they are eliminated from consideration. Fourth, almost all evidence comes from studies of children and adolescents with some but much less evidence to guide practice. Thus, some who work with adults may dismiss the evidence from children and adolescents as being irrelevant. Fifth, different professionals and researchers define the same problem in different way (Sturmey, 2009b). Why did this client throw the chair across the room? A behaviorist would say she emitted aggressive operant behavior that in the past had been reinforced by removal of irritating people. A cognitive behavioral therapist might say she has inappropriate beliefs and her labeling of the situation as threatening caused her to throw the chair. An SIT therapist might say she was hypersensitive to noise, and a psychiatrist might say she has an underlying depression. Hence, each ideological camp may dismiss evidence from each other’s perspective, because it does not address the “real” problem: The cognitive therapist can dismiss the evidence for behavioral treatment because it treats aggression, not anger; the psychiatrist can dismiss it because it does not treat the real problem of depression; etc. Finally, professional training in ID is often of poor quality. It barely prepares or fails to prepare new professionals on how to treat people with disabilities effectively and is often ambiguous about the status of EBP. Hence, new professionals are left at sea as how to acquire critical skills. In the process, many fail to learn important skills or learn ineffective or harmful ones.

A Challenge Revisited

At the beginning of this chapter, we considered an example of challenges to practitioners and services. The example of food refusal could have gone differently for the individual concerned. The educators could have been trained effectively in special education to use EBPs before becoming special education teachers. Given that food refusal is a relatively rare problem, it may be unlikely that they would have been skilled in that specific problem. If they were well trained in EBP, however, they could have applied functional assessment to the problem and derived an effective treatment from first principles. Alternatively, they could have recognized early on that they did not have expertise in this problem and could have sought effective consultation and training. Perhaps their school could have given them better support and supervision to promote effective professional skills to ensure they focused on child outcomes and being open to using outside consultants effectively, rather than defending ineffective and somewhat harmful treatment. Such an approach would have achieved child outcomes early and avoided harm to the child, answering the ethical imperative for EBP, and precious resources would not have been invested in ineffective treatment, answering the economic imperative for EBP.

References

Allen, L. A., & Woolfolk, R. L. (2012). Somatoform and factitious disorders. In P. Sturmey & M. Hersen (Eds.), *Handbook of evidence-based practice in clinical psychology: Vol. 2. Adults* (pp. 365–394). Hoboken, NJ: Wiley.

American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct*. Washington, DC: Author.

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *The American Psychologist, 61*, 271–285.

Applied Behavior Analysis International. (2010). *Statement on restraint and seclusion, 2010*. Retrieved August 16, 2013, from http://www.abainternational.org/abai/policies-and-positions/restraint-and-seclusion,-2010.aspx (accessed on October 11, 2013).

Bettelheim, B. (1967). *The empty fortress: Infantile autism and the birth of the self*. New York: The Free Press.

Bishop-Fitzpatrick, L., Minshew, N. J., & Eack, S. M. (2013). A systematic review of psychosocial interventions for adults with autism spectrum disorders. *Journal of Autism and Developmental Disabilities, 43*, 687–694.

Buysse, V., & Wesley, P. W. (2006). *Evidence-based practice in the early childhood field*. Washington, DC: Zero to Three.

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist, 24*, 409–429.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*, 7–18.

Cheng, M. (2012). *French autistic kids mostly get psychotherapy*. Retrieved August 12, 2013, from http://usatoday30.usatoday.com/news/health/story/2012-05-18/France-autism-psychotherapy/55056318/1 (accessed on October 11, 2013).

Chiesa, M. (1994). *Radical behaviorism: The philosophy and the science*. Boston: Authors Cooperative.

Chowdhury, M., & Benson, B. A. (2010). Use of differential reinforcement to reduce behavior problems in adults with intellectual disabilities: A methodological review. *Research in Developmental Disabilities, 32*, 383–394.

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry, 23*, 318–327.

Cuijpers, P., Van Straten, A., Driessen, E., Van Oppen, P., Bockting, C., & Anderson, G. (2012). Depression and dysthymic disorders. In P. Sturmey & M. Hersen (Eds.), *Handbook of evidence-based practice in clinical psychology: Vol. 2. Adults* (pp. 243–284). Hoboken, NJ: Wiley.

Duggan, L. M., Morris, M., & Adams, C. M. (1997). Prevalence study of the randomized controlled trials in the *Journal of Intellectual Disability Research*: 1957–1994. *Journal of Intellectual Disability Research, 41*, 232–237.

Dunst, C. J., Trivette, C. M., & Cutspec, P. A. (2002). *Toward an operational definition of evidence-based practices*. Ashville, NC: Winterberry Press.

Emerson, E. (2006). The need for credible evidence: Comments on 'On some recent claims for the efficacy of cognitive therapy for people with Intellectual Disabilities'. *Journal of Applied Research in Intellectual Disabilities, 19*, 121–123.

Eysenck, H. J. (1957). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology, 16*, 319–324.

Foa, E. B., Gillihan, S. J., & Bryant, R. A. (2013). Challenges and successes in dissemination of evidence-based treatments for posttraumatic stress: Lessons learned from prolonged exposure therapy. *Psychological Science in the Public Interest, 14*, 65–111.

Glass, G. V. (2000). *Meta-analysis at 25*. Retrieved August 19, 2013, from http://www.gvglass.info/papers/meta25.html (accessed on October 11, 2013).

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., et al. (2010). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal, 336*, 924–926.

Hagopian, L. P., Rooker, G. W., & Rolider, N. U. (2011). Identifying empirically supported treatments for pica in individuals with intellectual disabilities. *Research in Developmental Disabilities, 32*, 2114–2120.

Hamelin, J., & Sturmey, P. (2013). Psychological treatment: What does the evidence tell us? To appear. In E. Tsakanikos & J. McCarthy (Eds.), *Handbook of psychopathology in intellectual disabilities*. New York: Springer.

Hastings, R. P., Hatton, C., Lindsay, W. R., & Taylor, J. L. (2013). Psychological therapies for adults with intellectual disabilities: Future directions for research and practice. In J. L. Taylor, W. R. Lindsay, R. P. Hastings, & C. Hatton (Eds.), *Psychological therapies for adults with intellectual disabilities* (pp. 266–276). Chichester, UK: Wiley.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165–179.

Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academy Press.

Kazdin, A. E. (2010). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York: Oxford University Press.

Keenan, M., & Dillenberger, K. (2011). When all you have is a hammer…: RCTs and hegemony in science. *Research on Autism Spectrum Disorders, 5*, 1–13.

King, R. (2005) Proceeding with compassion while awaiting the evidence. Psychotherapy and individuals with mental retardation. *Mental Retardation, 43*, 448–450.

Knapp, M., & McDaid, D. (2012). Economics of evidence-based practice and mental health. In P. Sturmey & M. Hersen (Eds.), *Handbook of evidence-based practice in clinical psychology: Vol. 1. Children and adolescents* (pp. 71–94). Hoboken, NJ: Wiley.

Layard, R., Clark, D., Knapp, M., & Mayraz, G. (2006). Cost-benefit analysis of psychological therapy. *National Institute Economic Review, 202*, 90–98.

Lilienfield, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science, 2*, 53–70.

Maginnis, K. (2008). *Independent review of autism services*. Retrieved August 19, 2013, from http://www.dhsspsni.gov.uk/independent_review_of_autism_services_final_report.pdf (accessed on October 11, 2013).

McAdam, D., Breidbord, J., Dahl, A., & Williams, D. E. (2012). Pica. In M. Hersen & P. Sturmey (Eds.), *Handbook of evidence-based practice in clinical psychology: Vol. 1. Children and adolescents (pp. 323–338)*. New York: Wiley.

McAdam, D. B., Sherman, J. A., Sheldon, J. B., & Napolitano, D. A. (2004). Behavioral interventions to reduce the pica of persons with developmental disabilities. *Behavior Modification, 28*, 45–72.

Moher, D., Schulz, K. F., Altman, D. G., & the CONSORT Group. (2010). CONSORT 2010. Statement: Updated guidelines for reporting parallel group randomised trials. *BMC Medicine, 8*, 18.

National Autism Center. (2009). *National standards project*. Randolph, MA: Author.

National Institute for Clinical Excellence. (2012). *Autism. The NICE guideline on recognition, referral diagnosis and management of adults on the autism spectrum*. Leicester, UK: The British Psychological Society and The Royal College of Psychiatrists.

Palmen, A., Didden, R., & Lang, R. (2010). A systematic review of behavioral intervention research on adaptive skill building in high-functioning young adults with autism spectrum disorder. *Research in Autism Spectrum Disorders, 6*, 602–617.

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology, 31*, 109–118.

Rusch, A. J., & Frances, A. (Eds.). (2000). Expert consensus guidelines for treatment of psychiatric and behavioral problems in mental retardation. *American Journal on Mental Retardation, 105*, 159–228.

Sackett, D. L., Richardson, S. W., Rosenberg, W. R., & Haynes, R. B. (1997). Evidence-based medicine; how to practice and teach EBM. London: Churchill- Livingstone.

Schlosser, R. W., & Raghavendra, P. (2004). Evidence based practice in augmentative and alternative communication. *Augmentative and Alternative Communication, 20*, 1–21.

Schlosser, R. W., Wendt, O., & Sigafoos, J. (2007). Not all systematic reviews are created equal: Considerations for appraisal. *Evidence-Based Communication Assessment and Intervention, 1*, 138–150.

Schofield, H. (2012). *France's autism treatment 'shame'*. Retrieved August 12, 2013, from http://www.bbc.co.uk/news/magazine-17583123 (accessed on October 11, 2013).

Schreck, K. A., & Millar, V. A. (2010). How to behave ethically in a world of fads. *Behavioral; Interventions, 25*, 307–324.

Scotti, J. R., Evans, I. M., Meyer, L. H., & Walker, P. (1991). A meta-analysis of intervention research with problem behavior: Treatment validity and standards of practice. *American Journal on Mental Retardation, 96*, 233–256.

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York: Basic Books.

Skinner, B. F. (1953). *Science and human behavior*. New York: Pearson.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*, 372–360.

Sneed, J. R., Ferteck, E. A., Kanellopoulos, D., & Culag-Reinlieb, M. E. (2012). Borderline personality disorder. In P. Sturmey & M. Hersen (Eds.), *Handbook of evidence-based practice in clinical psychology: Vol. 2. Adults* (pp. 507–530). Hoboken, NJ: Wiley.

Straus, S., Glasziou, P., Richardson, W. S., & Haynes, R. B. (2011). *Evidence-based medicine. How to practice it* (4th ed.). London: Churchill Livingston.

Sturmey, P. (2006a). Against psychotherapy with people who have mental retardation: In response to the responses. *Mental Retardation, 44*, 71–74.

Sturmey, P. (2006b). On some recent claims for the efficacy of cognitive therapy for people with intellectual disabilities. *Journal of Applied Research in Intellectual Disabilities, 19*, 109–118.

Sturmey, P. (2006c). In response to Lindsay and Emerson. *Journal of Applied Research in Intellectual Disabilities, 19*, 125–129.

Sturmey, P. (2009a). Behavioral activation is an evidence-based treatment of depression. *Behavior Modification, 33*, 818–829.

Sturmey, P. (2009b). *Varieties of case formulation*. Chichester, UK: Wiley.

Travis, R., & Sturmey, P. (2013). Using behavioural skills training to treat aggression in adults with mild intellectual disability in a forensic setting. *Journal of Applied Research in Intellectual Disabilities, 26*, 481–488.

Van Houten, R., Axelrod, S., Bailey, J. S., Favell, J. E., Foxx, R. M., Iwata, B. A., et al. (1988). The right to effective behavioral treatment. *Journal of Applied Behavior Analysis, 11*, 111–114.

Ward-Horner, J. C., & Sturmey, P. (2012). Component analysis of behavior skills training in functional analysis. *Behavioral Interventions, 27*, 75–92.

Willner, P., Rose, J., Jahoda, A., Stenfert Kroese, B., Felce, D., Cohen, D., et al. (2013). Group-based cognitive-behavioural anger management for people with mild to moderate intellectual disabilities: Cluster randomized controlled trial. *British Journal of Psychiatry, 203*, 288–296.