

# BEYOND “ESTs”

## *Problematic Assumptions in the Pursuit of Evidence-Based Practice*

Paul L. Wachtel, PhD

*City College of New York and The Graduate Center, City University of New York*

There has been much confusion in the literature of psychotherapy between the broad concept of evidence-based practice and the narrower set of criteria that have been employed in designating certain treatments as “empirically validated” or “empirically supported.” In contrast to the appropriate concern with examining the evidence for the efficacy of various approaches to therapy and for the theoretical assumptions that underlie them, the “empirically supported treatments” movement has been characterized more by ideology and faulty assumptions than by good science. This paper examines in detail the scientific and logical limitations of the “EST” movement and aims to place the empirical investigation of theory and practice in psychotherapy on a sounder basis.

*Keywords:* Empirically Supported Treatments, Evidence-Based Practice, Psychotherapy Outcome, Psychotherapy Research, ESTs

Increasingly, in recent years, there have been calls for establishing the practice of psychotherapy on an evidence-based foundation. In principle, this is a salutary development. Unfortunately, however, the “empirically supported treatments” (EST) movement, which has largely dominated discussion of evidence-based practice in recent years, has been characterized by a set of assumptions that impede sound understanding of the sources of therapeutic change and generate biased conclusions regarding what therapeutic approaches are actually helpful to patients. I aim in this paper to examine closely these “EST” assumptions and to indicate an alternative view of how clinical practice can be rooted in respect for evidence.

The reader will notice that I have placed the terms “empirically supported” and “EST” in quotation marks. I do so throughout this article because I do not wish to further contribute to the misconceptions that result when the concepts of empirical validation or empirical support are ceded to the advocates of a particular tendentious

---

Paul L. Wachtel, PhD, Department of Psychology, City College of New York, and The Graduate Center, City University of New York.

Correspondence concerning this article should be addressed to Paul L. Wachtel, PhD, Department of Psychology, City College of New York, 138th Street and Convent Avenue, New York, NY 10031. E-mail: paul.wachtel@gmail.com

definition of those ideas. It reflects a problematic acceptance of faulty premises when critics of this parochial methodology say things like “it is important that training programs teach therapeutic approaches other than ESTs as well.” Such statements seem to accept the idea that “ESTs” are the only therapies that are empirically supported, and then try to battle for some space for therapies that are *not* “ESTs,” as if those other therapies, though not empirically supported, have some *other* virtue. In fact, as I shall argue in this article, there are serious flaws in the empirical support for many “ESTs” as therapies applicable to the majority of patients who seek therapy and, conversely, there is often evidence at least as strong or stronger supporting therapeutic approaches *not* on the “EST” lists that have been promulgated (see, e.g., Shedler, 2010).

### What Constitutes Empirical Validation? Shifting Sands and Mutating Rhetoric

One good indicator of the conceptual confusion and ideological scrambling that has characterized much of the literature on the empirical foundations of therapeutic practice is the shifting vocabulary that has characterized the debate. In 1995, the Task Force on Promotion and Dissemination of Psychological Procedures, a group originating in the clinical psychology division of the American Psychological Association (APA), published a list of treatments that were deemed to be “well-established” or “probably efficacious” (Task Force on Promotion and Dissemination of Psychological Procedures, 1995). In the relatively few years since this list appeared, the *nom de guerre* of the movement created by these activists has mutated with some regularity. The first shift in the rhetoric was from “well established” or “probably efficacious” to “empirically validated.” Before long, however, this terminology too gave way, under pressure from critics who noted that it was not consistent with a genuinely scientific attitude to claim that the approaches on the list had been “validated” when the cumulative findings of research over time so often lead us to modify our initial enthusiasms. Thus, a new version of the list then appeared—though with little change in the criteria for inclusion or exclusion—under the name “empirically *supported*” treatments. Presently, that terminology too is in the process of being jettisoned, with “evidence-based practice” the rhetoric du jour.

It is also instructive to note that the names of the task forces themselves also showed a persistent tendency to morph. The first task force created by Division 12 of APA was called the Task Force on Promotion and Dissemination of Psychological Procedures. Before long it had become the Task Force on Psychological Interventions. Not long after, we had the Standing Committee on Science and Practice (Sanderson, 2003). These names, it is worth noting, moved progressively away from explicitly revealing their interest in “dissemination” and toward the kind of “neutral” sounding names that characterize oil industry advocacy in terms such as “Committee on Energy Policy.” Even the term “task force,” with its clear trace of militancy, has now been retired, replaced by the blander term committee. Yet even a cursory look reveals that the aim of disseminating to the general public the group’s particular point of view on what has and has not been validated has remained unmodified. (See below for further discussion of this aim and the actions that the task forces and committees have taken.)

Such rapid shifts in terminology, with little if any real modification in the essential idea, suggest some unacknowledged sense of unease with the arguments being made, an unease that is periodically covered over by strategic rebranding. There was, I wish to suggest, good reason for unease; for beneath these varying cosmetic makeovers lay a

single, rather fixed set of criteria by which judgment was made regarding whether a particular therapeutic approach was “well established,” “probably efficacious,” “empirically validated,” “empirically supported,” or “evidence based.” These criteria, moreover, were remarkably tendentious, a set of standards that constituted an ideological litmus test much more than any genuine requirement of adherence to the scientific method. Put differently, the problem was not, as was not so subtly implied, that the various task forces that advocated for these shifting labels and fixed criteria were attempting to impose scientific discipline on an unwilling, antiscientific community of practitioners; it was that the recommendations of these task forces were *not sufficiently* respectful of the canons of science. The positions taken by these task forces introduced criteria that, on close inspection, not only did not represent very good science but actually *impeded* good science.

In considering where the criteria advocated by the various task forces and committees fall considerably short of adequate science, there are at least four features that require our attention. I will discuss in succession the emphasis on patients in a study being limited to a single diagnostic category, on manualization, and on randomized, controlled trials (RCTs). After explicating the ways in which each of these three criteria can be misused to both restrict and misrepresent the available evidence, I will then turn to a fourth characteristic of the movement I have been discussing—the dichotomous thinking that leads to the promulgation of lists of treatments that are empirically validated and to a clearly implied shadow list of those that purportedly are not.

### People or Disorders?

One of the criteria that has been central in the generation of the various lists of “well-established,” “probably efficacious,” “empirically validated,” “empirically supported,” or “evidence-based” treatments has been limiting the focus of the therapeutic effort to a single, well-defined problem. The rationale for this requirement at first sounds deceptively reasonable. As generally stated, the point is simply that “Characteristics of samples must be specified” (e.g., Chambless & Ollendick, 2001, p. 689). Who could object to that? Don’t we want to know who the patients were and how generalizable the findings are? Don’t we want precision in our formulations? At least since Kiesler’s (1966) classic critique of psychotherapy research to that point, concern about *what works for whom* (Roth & Fonagy, 2004) has long been a mark of clear thinking. The problem arises when we look more closely at what this actually means in the specific application of compiling lists of “empirically supported” therapies. Virtually without exception, what is really meant by specifying the characteristics of the sample is limiting the study to *a particular Axis I category in the then-current DSM*. This requirement has increasingly been adopted by granting agencies, who similarly insist on such sample limitations.

As Westen, Novotny, and Thompson-Brenner (2004) have pointed out, this insistence effectively excludes attention to the treatment needs of a vast portion of the patients who seek psychotherapy. Most obviously excluded are the very large number of patients who have multiple complaints (so-called comorbidity), itself a far larger number of people than those who meet the “EST” paradigm’s criteria. But those criteria also exclude those people who, although suffering and seeking help, do not readily fit the criteria for *any* particular DSM diagnosis. The number of patients in this latter category alone has been estimated, based on the research conducted thus far, at *between one third and one half of all people*

*seeking mental health treatment* (Westen et al., 2004). Moreover, even for the relatively small number of people who do fit the criteria for one and only one DSM Axis I diagnosis, Westen et al. (2004) point out that, “In the modal efficacy study, somewhere between 40% and 70% of patients who present for treatment with symptoms of the disorder are excluded (not including the unknown percentage of patients who are screened by referring clinicians who know the study’s exclusion criteria).” Thus, taking all the exclusions into account, the typical “EST” criterion of restricting the study to a single Axis I disorder essentially guarantees that research on the efficacy of psychotherapy will be conducted on samples that differ very substantially from the majority of patients actually in psychotherapy. This is a rather strange way to extend the reach of science into daily clinical practice.<sup>1</sup>

It should also be noted that this restricting of “researchable” therapeutic work to patients with a specific DSM diagnosis fails to consider that for a large portion of the people who enter therapists’ offices, defining the problem *is* to a very large extent the problem; or put differently, for these patients the therapist’s most important skill or contribution lies in helping the patient identify and articulate just what is bothering him. Relatedly, Kazdin (2008) notes that even when symptoms are the focus of treatment, “over half of patients seen in therapy add new target complaints or change their complaints over the course of treatment” (p. 148).

### Manuals and Rituals

A second feature of the standard “EST” criteria is perhaps even more problematic—the requirement that the treatment be manualized. Here again, proponents of these criteria offer a reasonable sounding rationale—to evaluate whether the treatment being investigated in any particular study was effective, we need to know what the actual treatment was. But here again, what has emerged has been a tendentiously conceived and extraordinarily narrow investigative strategy. Instead of being viewed as one of a variety of possible solutions to the scientific challenge of specifying the actual therapy employed in a study, manualization has increasingly become a *requirement* by granting agencies for funding research, and training in “manualized treatment” has become widely (and falsely) equated with training in therapeutic approaches that are based upon solid and reliable evidence. Most problematically, the question-begging logic of the “EST” paradigm essentially implies that a nonmanualized therapy cannot *by definition* be empirically validated or supported since (with very limited and grudging exceptions) manualization has been treated as a fundamental requirement for empirical support per se. This is not a championing of science; it is an *abdication* of science, a decision *not to investigate* nonmanualized treatments that bespeaks at best a poverty of imagination in addressing methodological challenges.

It might be objected that the “EST” paradigm does not strictly require a manual. Chambless and Ollendick (2001), for example, in an influential statement of the “EST” approach, depict as the requirement “treatment manuals *or their equivalent* in the form of a clear description of the treatment.” Now, I am as in favor as they are of clear description,

---

<sup>1</sup> The degree of exclusion remains a topic of controversy. See, for example, Westen, Stirman, & DeRubeis, (2006). Less controversial, and at least as problematic, is that most of the research on which the “EST” lists have been based has been rather strikingly inattentive to issues of ethnic, racial, and other forms of diversity (e.g., Sue and Zane, 2006).

but this seemingly more open and reasonable statement is not consistent with the actual history of the "EST" movement, whose proponents have consistently dismissed an enormous body of evidence supporting the therapeutic impact of treatments other than those on the "EST" lists. This dismissive approach to uncongenial data was evident as early as the original Division 12 Task Force on Promotion and Dissemination of Psychological Procedures (1995). In their first published statement, attempting to discredit the influential review by Smith, Glass, and Miller (1980), which they noted had "convinced many that substantial evidence demonstrated the efficacy of psychosocial treatments," the task force publication stated, "Finally, and *perhaps most important*, the studies in the Smith et al. review predated the standardization of treatments in research studies *through the use of treatment manuals*" (p. 3, italics added). The two italicized phrases reveal the degree to which manuals were made the linchpin of an effort to prescribe one and only one methodology for psychotherapy outcome research and to *dismiss or ignore* evidence gathered in other ways, however careful, methodologically sophisticated, and appropriate to the problem at hand. The second Division 12 Task Force (Chambless et al., 1996) states quite explicitly that manualization was virtually a sine qua non for them to regard a treatment as empirically validated, with only "specific and *rare* exceptions" (p. 6, italics added).

My point is not that the creation of therapy manuals is never appropriate or useful. A considerable range of therapies (including some fairly complex psychodynamic and humanistic approaches) have been manualized for research purposes, and I do not mean to deride the efforts of these investigators. Rather my point is that to make manualization a *requirement* for regarding a treatment approach as evidence-based is not a reflection of commitment to scientific rigor, but a political ploy that effectively excludes from the lists of evidence-based treatments a variety of treatments for which there is in fact a very substantial body of evidence (see below), but which do not happen to have approached the task of empirical validation via the particular investigative strategies that the "EST" movement advocates.

A personal experience that predates the concept of manualized therapy helps to provide some perspective on the current preoccupation with manuals. In a study conducted with Jean Schimek (Wachtel & Schimek, 1970), we were attempting to trace the effects of emotionally toned incidental stimuli on subjects' mood, fantasies and thought processes. While subjects were creating Thematic Apperception Test (TAT) stories and participating in other measures, a tape recording was played in the next room in a way designed to sound like it was the sound of some activity going on next door that was not relevant to the study. In one condition, the sound was of pleasant conversation and laughter; in the other, it was of arguing. The tape recordings were played at a low volume so that no words could be made out, and no mention was made of them during the procedure. On careful and probing debriefing, only three of 60 subjects reported any suspicion at all that the incidental stimulus was part of the study.

In attempting to assess whether the two different incidental stimuli led to different degrees of aggression in the subjects' TAT stories, we at first attempted to spell out in great detail the criteria for scoring a protocol as having more or less aggression in it. In essence, we tried to create a scoring "manual." This effort proved frustrating in the extreme. It took an enormous amount of rather obsessional work, and in the end, for all our efforts, we could not get adequate interrater reliability. Finally, out of frustration and a growing concern that we were getting nowhere, we decided to see what would happen if, *without* a "manual," the stories were simply rated on the basis of "how angry they

seemed.” That is, the switch was made from following a set of detailed, manualized instructions, to relying on what Polanyi (1958, 1967) called tacit knowledge, tapping a perception or understanding that may not yet be able to be explicitly articulated but may be reliably and accurately registered nonetheless. The result was surprising and heartening. When the task was approached in this *less* spelled out way, we achieved quite satisfactory reliability, and, moreover, the data were consistent with the study’s hypothesis: the subjects exposed to the aggressively toned incidental stimulus evidenced more aggression in their TAT cards. (Obviously, raters were blind as to which condition the subject was in.)

The study has interesting implications for understanding the role of implicit processing in the generation of affective experiences, and in this respect bears on the empirical foundations of psychoanalytic formulations as well as a range of other perspectives in contemporary personality and social psychology. But I mention it here for a different reason. What it demonstrates, in the present context, is an important *methodological* point; often, it is not necessary—and may even be counterproductive—to spell out in manual form the rules for discriminating between different psychological phenomena. In the realm of psychotherapy research, the implication of what I have just described is that the important and salutary aim of knowing just what kind of treatment one is evaluating in an outcome study is not necessarily best served by creating an elaborate manual; indeed, this aim may even be impeded by such an effort. If one wants to evaluate whether a purported psychodynamic treatment, or cognitive–behavioral treatment, or any other kind of treatment was in fact what was being practiced in the study, one of the best ways is often simply to ask practitioners of the approach being evaluated to watch or listen to a set of sessions and to rate, based on the complex criteria they hold in their head (again see Polanyi, 1958, 1967, on tacit knowledge) the degree to which each session matches the prototype. Such an approach renders “nonmanualized” treatments fully accessible to rigorous empirical assessment.

Even better perhaps than the approach I have just described is the employment, again without the treatment necessarily being manualized, of ratings based on measures such as the Psychotherapy Process Q-Sort [PQS] (Jones, 2000; see also Ablon & Jones, 1998; Jones & Pulos, 1993), an instrument designed not to detect the presence of the brand name therapy “packages” that are the focus of the “EST” approach (Ablon, Levy, & Katzenstein, 2006; Pachankis & Goldfried, 2007; Rosen & Davison, 2003; Westen et al., 2004), but rather of very specific kinds of comments and behaviors. It is ironic that advocates of the “EST” approach, who have advocated restricting the patient sample to a single Axis I diagnosis under the banner of precision and specificity, place such enormous emphasis on anointing rather global packages of interventions, which, when closely examined, often represent a hodgepodge of actual elements and interventions (Ablon & Jones, 2002; Castonguay, Goldfried, Wisner, Raue, Hayes, 1996; Shedler, 2010; Westen et al., 2004). In contrast, the PQS approach addresses itself to the specific processes and interventions that account for therapeutic success.

Further considering the implications of attempting to “manualize” complex forms of treatment, it is important to be clear that each application of the manual is a new application of the general principles around which the manual is organized, following some general (and even reasonably specifically detailed) guidelines, but applying them in a somewhat different way with each patient. This is the element of truth in the complaint that “EST” advocates make about their critics’ characterizations of manualized treatments. On the Web site of the Association for Behavioral and Cognitive Therapies, for example, under the heading titled “ESTs: Misconceptions and Misunderstandings,” they list as a

misconception that “Because ESTs are manualized, they necessarily constrain clinical creativity,” and then further respond as follows:

To some extent, this criticism is based on a caricature of manualized therapies. Treatment manuals do not necessarily mandate fixed responses to patients’ verbal behaviors in therapy; instead, most manuals provide flexible guidelines for how to proceed at different stages of treatment. Moreover, increasing numbers of treatment manuals afford therapists considerable leeway to respond flexibly to differing patient trajectories within treatment.<sup>2</sup>

For a subset of manualized treatments, this is an accurate account, and it reflects an approach of benefit to many patients. As Beutler, Castonguay, and Follette (2006) point out, the available research strongly indicates that “therapist flexibility in changing strategies, adapting to patient presentations, tolerance, and creativity are related to improvement.” But the very fact of this degree of flexibility means that what is assessed in RCTs of such treatments is the effectiveness, essentially, of the *principles* or the *basic strategy* on which they are grounded not the effectiveness of a specific procedure per se. (See the discussion later in this article of the differences between a research program focused on *principles* and one addressed to treatment “packages.”) If one looks closely at what is actually observed in these studies, the question really addressed is: On average, do the *various versions* of what is laid out in general form in Manual X, prove to be more helpful than some comparison group?”

This is not an unimportant or uninteresting question. But it *is* a question that is often not approached in an honest or clear-eyed way. It is approached as if “the treatment” is the same for each patient, as the pill is in a pharmaceutical RCT or the experimental condition is in a tightly controlled laboratory experiment. As Westen et al. (2004) point out, even in the brief 12, 16, or 20 session therapies investigated in most “EST” style studies, the threats to internal validity accumulate rapidly over the period of the study. Comparing therapy RCTs with the typical 15- or 30-min psychology experiment, they note wryly that, “we are aware of no other experiments in the history of psychology in which a manipulation intended to constitute a single experimental condition approached that length” (p. 633).

The problem is not so great if all one wishes to demonstrate is that *something* in the designated treatment condition made it work better than the alternative treatments (or placebo or no-treatment conditions). If the sample is large enough and the randomization is genuine, this much still seems solid. But the claim that *manualizing* the treatment is essential for understanding just what kind of treatment was actually administered is faulty, a ploy rather than a serious methodological necessity. It is *not* in fact the manual that determines what kind of therapy is being practiced by the therapists in the study. One can have a well-designed and perfectly clear manual that is not being skillfully or faithfully followed by the therapists in the study. Ultimately, the criterion for whether the kind of therapy thought to be under investigation actually was the therapy being practiced is not the manual but the *adherence checks* that are conducted. And, as the incidental stimulus study described earlier illustrates, an adherence judge who is equipped intellectually and emotionally to assess what is being evaluated can often do as well—and sometimes even better—*without* a manual as with one. With a manual that allots substantial flexibility to the clinician to respond to the patient’s individuality and experience of what is transpiring—the very case the Association for Behavioral and Cognitive Therapy (ABCT) Web

---

<sup>2</sup> <http://www.abct.org/dMental/?m=mMental&fa=MythsAndFacts>

site makes in refuting what it complains are misrepresentations—“the manual” is very largely a facade. It provides little validation that is not equally well conveyed by adherence checks or prototype matching. It does, however, serve as a formalistic barrier to ensure that “unmanualized” treatments will not be listed as “empirically supported.”

### Randomized Controlled Trials

The third problematic criterion in the “EST” paradigm is the insistence on RCTs. The difficulties with this third criterion may be harder to see initially and may require still closer examination, because the rationale for emphasizing RCTs is, on the face of it, considerably stronger than for the other two problematic criteria just discussed. For whereas the use of a manual or an exclusive focus on a single DSM diagnosis does not in itself offer much real gain in confidence in the findings, the employment of RCTs does. Where they are feasible and appropriate, RCTs do offer a degree of control over potential confounding variables that is of genuine and significant evidential value. But the claim that RCTs are the “gold standard”—and the implicit message that we should regard as of only very limited evidential value data obtained in any other way—is based on a rather simplistic understanding of scientific inquiry and of the range of ways to examine competing hypotheses and explanations.

The idea that evidence obtained via methods other than RCTs is suspect—indeed so suspect that any therapy not investigated in this particular manner is not eligible to be regarded as empirically supported or evidence based—reflects a view of science that would warm the hearts of Creationists and other opponents of Darwinian theory. For virtually all of the vast body of research that supports evolutionary theory is not research of the sort that “EST” advocates depict as the “gold standard.” If no hypothesis that has not been tested by randomized assignment to experimental and control conditions should be viewed as validated, evolutionary biology would be the “mere theory” that the Creationists portray it as. For although there certainly have been experimental studies that have addressed various implications of Darwinian theory, by and large the evidence for evolution rests on nonexperimental evidence, as does most of the evidence in sciences such as geology or cosmology. When the subject of a science addresses processes occurring over millions of years, not to mention processes we are simply incapable of controlling, the evidence obviously cannot be addressed via “random assignment.” This does not, however, render these sciences dependent on mere impressions or opinions. A wide range of sophisticated quantitative methodologies and apt choices of observations to serve as controls has enabled these sciences to develop an empirical foundation and a degree of methodological rigor to which our own field can still only aspire.

The *real* “gold standard” in science is respect for the evidence, wherever it leads us, and the employment of the best and most appropriate methodology *for the particular problem being studied*. If one cannot employ one’s favorite methods to study a phenomenon or a theoretical claim, the fault does not lie in the phenomenon; it lies in the methods. Goldfried and Eubanks-Carter (2004) have illustrated this point with a charming anecdote:

The mathematician and astronomer Sir Arthur Eddington (1939), in highlighting some of the limitations inherent in scientific research, told the story of a scientist who wants to study the size of the fish in the sea. To do so, the scientist weaves a fishnet with a 2-inch mesh and collects systematic and representative samples in different locations, under different weather conditions, over different periods of time. The scientist carefully measures and classifies the

fish in each catch, and draws the following conclusion: There are no fish in the sea under two inches (Goldfried & Eubanks-Carter, 2004, p. 669).

The lists of “ESTs” report the fish that happen to be caught by the particular methodology that EST advocates choose, and it is misleading and erroneous to conclude that those are the only efficacious fish in the therapeutic sea.

### *RCTs and the Absence of Double-Blind Studies*

In an interesting irony—because many of the leading “EST” advocates represent therapeutic orientations that originated as a challenge to the purported “medical model” of psychoanalysis—“EST” advocates insist that all valid conclusions about the efficacy of any particular therapeutic approach must rest on a methodology that essentially mimics the structure of drug trials in medical research. However, they do so without considering sufficiently what makes the RCT methodology appropriate to that realm of investigation. The use of RCTs in drug trials almost always also includes, as an essential element, the employment of a double-blind methodology. Neither the patient nor the doctor administering the medication knows whether any particular patient is receiving the medication under investigation or a placebo coated to be identical in appearance. Indeed, when the side effects of the active medication are such that they are readily detected by either party, the internal validity of the study is seriously compromised. In contrast, in studies of psychotherapy, no one is unaware of which treatment is being offered or received. Without this crucial feature of the drug studies that the “EST” methodology attempts to mimic, the “gold standard” looks more like painted tin.

The absence of (and indeed, in most cases, the virtual impossibility of) a double-blind methodology in psychotherapy outcome studies is probably one important factor contributing to the finding by Luborsky et al. (1999) that most studies end up demonstrating the superiority of whatever approach the investigator is most closely allied to. The knowledge, not only by the investigator but by the therapist, of which procedure is being administered introduces powerful—and impossible to measure—influences on what actually transpires in the room when the therapist is practicing one approach or another.

This methodological limitation does not render psychotherapy outcome research meaningless, but it does put into perspective the “gold standard” rhetoric that has been employed by “EST” advocates and places the findings cited by them on a much more even footing with the large number of studies that support the efficacy of approaches *not* on the “EST” lists (see, e.g., Hubble, Duncan, & Miller, 1999; Lamb & Jones, 2009; Leichsenring & Rabung, 2008; Shedler, 2010; Wampold, 2008). Many of those studies utilized methodologies that, when the considerations I have been advancing in this article are taken into account, have every bit as much evidential value as the studies on which the “EST” lists are based. In fact, many of the studies utilized RCTs as well, though often without the ritualistic manual or limited target population.

### *Multimethod Approaches to Studying the Process and Outcome of Psychotherapy: The Strengths of RCTs and the Complementary Strengths of Alternative Methodologies*

Results that seem persuasive utilizing one methodology not infrequently fail to hold up when a different measure or a different methodology is used. A finding which we thought to be reflective of a general principle may turn out to be specific to, or an artifact of, the particular measures or situations employed or sampled. In the realm of assessing person-

ality traits, it was essentially this understanding that led Campbell and Fisk (1959), in a classic contribution to research methodology, to advocate the use of multimethod-multitrait approaches. But much the same logic points to the necessity of utilizing complementary methodologies in studying the process and outcome of psychotherapy as well.

RCTs clearly offer a powerful methodology for drawing inferences about what kinds of therapeutic interventions actually help people. As one part of a broader research effort, they are a tool of noteworthy evidential power. But the “best” methodology cannot be determined in the abstract; in any scientific work, we must find the best available methodology *for the problem at hand*. Every graduate student learns in his or her first statistics or methodology course about the limits of inferring causality from correlations and about the power of experimental research designs that randomly assign subjects to one or another condition. But it is important not to place more weight than it can bear on this rather general and rudimentary account of a much more complex state of affairs. The “EST” guidelines would have us build our science on a precariously narrow foundation and, moreover, a foundation that is based on ignoring relevant and important data. Although it remains correct that one cannot *with certainty* infer causality from correlation, there have evolved over the years a host of sophisticated correlational designs (such as, with some overlap in the terminology, time series analyses, path analysis, hierarchical regression analysis, multidimensional scaling, structural equation modeling, latent class modeling, and cross-lagged panel designs) that render the gap in evidential value between experimental and correlational designs far less absolute than was once taught in Methodology I courses.

Moreover, as the concept of Type I and Type II error was developed to represent, there is no free lunch in the evaluation of research findings. We can err in accepting as valid a finding that is actually attributable to chance variations, but we can equally seriously err in rejecting an important and valid phenomenon that did not, in a particular study, achieve “statistical significance.” After all, when a difference is obtained that is significant “only” at, say, the .10 level, that means that the odds are 9 to 1 that it is a *real* difference.<sup>3</sup>

Similar tradeoffs to those in the realm of statistical inference are evident in the realm of methodology. When correlational and naturalistic quasi-experimental data are summarily dismissed by “EST” advocates as not definitively probative, useful knowledge is not ensured. There is an equivalent to Type II error in this realm as well. Insisting on only randomized controlled trials (even apart from the tendentious criteria the “EST” paradigm *requires* of those trials—manuals, single Axis I diagnosis, etc.) virtually guarantees that much valuable data on real and useful contributions to therapeutic gain will be ignored, and hence erroneous conclusions drawn. There is a price to pay (both in knowledge and human welfare) not only in continuing to practice in ways that are not actually valid but in *failing* to recognize the value of practices whose efficacy one’s methods have not been able to demonstrate. There is no automatic solution to this dilemma. To reach an intelligent conclusion, one must weigh all of the evidence and all of the considerations, not just rely on a single metric, falsely regarded as definitive or a “gold standard.” When one sits complacently with a “no evidence” conclusion based on the absence of a “gold

---

<sup>3</sup> The limitations of considering statistical significance alone are one of the reasons that attention to effect sizes has increasingly been emphasized by serious researchers. The very same study that demonstrates that an effect is very likely to be “real,” rather than just a chance variation, may also demonstrate that it is a *very small* effect. I will have more to say about this shortly.

standard” demonstration of causal efficacy via an RCT, while a host of merely “probable” indications pile up, none of them in themselves absolutely definitive about the causal relationship between intervention and outcome, but the sum of them amounting to a far smaller likelihood of pointing in the wrong direction than one or two RCTs—which, of course, *also* are by no means absolute—one is engaging in a sophomoric caricature of science.

There are always a host of alternative explanations for why a particular therapeutic protocol in an RCT may achieve a measurable change, and some of them may be quite disparate from what the researcher thinks was the operative causal factor (see, e.g., Ablon & Jones, 1998, 2002; Castonguay et al., 1996; Jones & Pulos, 1993; Shedler, 2010). It is precisely in the effort to understand *why* particular therapeutic protocols are effective—in investigating the mediating and moderating variables (Kazdin, 2007, 2008) and the operative principles and processes—that we are enabled not just to *evaluate* our effectiveness but to *improve* it.<sup>4</sup> And it is here most of all that the “EST” movement’s “horse race” approach to therapy research lets us down.

This state of affairs might not matter so much if the treatments focused on by the “EST” proponents were particularly powerful treatments. But often they are not. In many cases, though in certain respects they may be said to “work” (that is, to yield statistically greater changes than some other prepackaged treatment or than no treatment at all), the very studies on which “EST” advocates base their lists are often good evidence for their being *weak* treatments. A number of writers have pointed out (e.g., Kazdin, 2006, 2008; Westen et al., 2004) that in many of the studies, patients may evidence changes that are *statistically* significant but not necessarily *clinically* significant. The modal patient remains symptomatic (if *less* symptomatic) and often still meets the criteria for the diagnosis he held when the treatment began. Moreover, in many of these studies, long-term follow-up is limited or lacking, and what follow-up there is indicates that there is a good deal of relapse and that a significant portion of the patients who were treated with the supposedly “empirically supported” treatment sought further therapy elsewhere after completing the treatment (see, e.g., Westen et al., 2004; Westen & Morrison, 2001; Shedler, 2010). There is thus *very strong* reason for seeking a deeper understanding of these treatments so that their often mediocre results can be improved. But the “EST” approach, with its almost exclusive emphasis on RCTs and its inattention to process-outcome studies and other naturalistic and quasi-experimental designs that can better illuminate when and why the therapy succeeds or fails, presents an impediment to that effort.

### Misleading Lists and Deceptive Communications

A central preoccupation of the “EST” movement has been the creation of lists of therapies that have purportedly been shown to be empirically supported. The list makers do regularly acknowledge that their lists are not intended to imply that any therapy *not* on the list has been shown to be ineffective. They rest content with the (often equally false) claim that they have simply not been shown to *be* effective (as, supposedly, those on the list *have*

---

<sup>4</sup> Some attention to these matters is reflected in the “dismantling” research that sometimes follows up on the “winners and losers” emphasis that is primary in the “EST” approach. But this represents only a limited approximation to the serious and intensive examination of underlying processes and principles (see the next section of this article).

been). There are at least two major flaws in their contentions. First, even if they literally have not said that those not on the list are not effective, this is clearly how many people will read what they are saying. Second, there is in fact *a great deal* of evidence supporting the efficacy of a significant range of therapies not on the “EST” lists (see below).

It is certainly true that creating a list of, say, “probably efficacious” treatments doesn’t *logically* imply that therapies not on the list are “probably *not* efficacious;” but *psychologically*, such statements clearly do imply such a message. The neural circuits evoked by the statement “X has been proved to be effective but Y has not” are likely to be very similar to those evoked by “X is effective and Y is ineffective.” To pretend this is not the case is disingenuous. Any psychologist even reasonably familiar with the social psychological literature on persuasion and influence would be very skeptical of the proposition that one can offer a list of therapies that have been labeled as “empirically validated” and not influence people to think the others must be invalid.

Indeed, it is difficult to believe that such a message was not the real intent of the “EST” movement. The Division 12 Task Force on Promotion and Dissemination of Psychological Procedures (1995), for example, going well beyond simply presenting what they regarded as the evidence supporting certain approaches, recommended that “APA site visitors for accreditation of doctoral programs . . . make training in empirically validated treatments a high priority issue” (p. 7) and even advocated that “site visit teams make training in empirically validated treatments *a criterion for APA accreditation*” (p. 7, italics added).

Were the intent of such recommendations to ensure that students were trained in approaches that, based on the full range of valid methodologies, had achieved a reasonable level of empirical support, it might reflect a salutary respect for evidence and for the interests of the public. But from the context of the statement, it is clear that the intent of this proposal is *not* really to ensure that students are trained in methods that are empirically supported, but that they be trained in methods supported *by the “EST” movement’s criteria*, which is something quite different. In fact, to the degree that this view of training for psychotherapists is adopted (and troublingly, it is becoming the mode in many graduate schools), it will skew training in a direction that has problematic implications for public health. By focusing training efforts on brief manualized treatments for very narrowly defined complaints, it would skew training toward treatments designed to meet the mental health needs of only about 20% of the overall population seeking clinical assistance (see Westen et al., 2004).<sup>5</sup>

Turning to the second problem with the “EST” lists, even if one were to grant the improbable claim that creating lists of therapies that are “empirically supported” would not be perceived by readers of these lists as implying that those not on the list have been shown to be ineffective, the lists are still misleading. One might say that the judgment “no evidence for” is less severe than “evidence against,” but it is still a rather strongly

---

<sup>5</sup> Recall in this connection some of the limitations in the “EST” sample discussed earlier. First, a large number of people are excluded because they do not suffer from a single, neat Axis I diagnosis. This includes people with multiple symptomatic complaints (so-called co-morbidity); people whose difficulties are rooted in broad personality characteristics, including but not limited to people with Axis II diagnoses; people who do not appropriately fit into any DSM category, but who clearly suffer psychologically and are seeking relief from a therapist; patients, in most of the studies, who are suicidal or actively psychotic; and, finally, all those screened out by the referring sources who pick up that this is a person not appropriate for the study. Do the math – it’s mostly subtraction – and it is apparent that shifting training to “ESTs” would do a poor job of meeting the needs of the community.

invidious claim. Moreover, it is an inaccurate one. For many of the modes of treatment that are not on “the list,” the basis for that absence is ideological, not empirical. That is, as I have alluded to earlier, there is in fact *a great deal* of evidence for the efficacy of a significant range of therapies other than those on “the list” (see, e.g., Hubble et al., 1999; Levy & Ablon, 2008; Leichsenring, 2005; Leichsenring & Rabung, 2008; Shedler, 2010; Wampold, 2008). The studies providing this evidence are diverse and rigorous. But many of them do not incorporate the methodological dogmas of the “EST” movement. The therapies being investigated may not be manualized, or the study may not be restricted to a narrow segment of patients, and hence, even when the studies are rigorously conducted RCTs, they are ignored by “EST” advocates. But they provide robust evidence for the effectiveness of a far broader set of therapeutic approaches than “EST” advocates acknowledge. In fact, the evidence for the efficacy of psychodynamic treatments is not only very solid, but the evidence indicates as well that these treatments often have stronger and more enduring effects than the so-called “ESTs” do. As Shedler (2010) comments, “Especially noteworthy is the recurring finding that the benefits of psychodynamic psychotherapy not only endure but increase with time, a finding that has now emerged in at least five independent meta-analyses . . . In contrast, the benefits of other (nonpsychodynamic) empirically supported therapies *tend to decay over time* for the most common disorders” (italics added).

Especially problematic in the “EST list” approach to evidence-based practice is that “EST” advocates have not confined their arguments to the professional and research communities that have the training to evaluate them, but have attempted to influence the general public. On the Web site of the division of clinical psychology of APA—a division of which I have long been, and continue to be, both a member and a fellow, but which has seemed to me, over the years, increasingly to represent only a particular ideological subset of clinical psychologists—there is a section that is clearly directed to the lay public.<sup>6</sup> The web page is labeled “A Guide to Beneficial Psychotherapy,” with a following tag-line of “Empirically Supported Treatments.” The language reveals its focus on the lay reader and on potential psychotherapy patients quite clearly:

What approaches to psychotherapy are beneficial? Is my psychotherapy helping me with my problems? Is the psychotherapy I am considering likely to be beneficial for me?

These questions are direct and straightforward ones, and psychotherapy patients, and those considering becoming patients, deserve direct and straightforward answers. Unfortunately, these answers are not always easy to obtain.

To address consumers’ needs for information about benefits of psychotherapy, this Web site has been developed by the Committee on Science and Practice of the Society of Clinical Psychology, a division of the American Psychological Association, to provide brief descriptions of various psychotherapies that have met basic scientific standards for effectiveness.

---

<sup>6</sup> The material listed here was obtained from the Web site in November 2009 under the web address [http://www.apa.org/divisions/div12/rev\\_est/index.html](http://www.apa.org/divisions/div12/rev_est/index.html). When I looked at the Web site in December 2009 to check on a formatting issue, I could not find the same page. Both the address and the content had changed. One could say that this represents the process of continual refinement that comes with being scientifically and ethically responsible. One could also interpret the change as reflecting the same less salutary considerations that I discussed earlier in this article regarding the continually morphing names for the task forces and committees that compiled and promulgated these lists and the similarly shifting names for the definitions of the lists themselves. Be that as it may, it remains the case that in late 2009 the web page had contained the exact content I quote here.

It would hardly be unreasonable for the unknowledgeable lay reader to conclude that any other therapeutic approach has not met *even basic* scientific standards. Certainly they would be unlikely to know that “basic” here means the standards of a particular subset of psychologists with a particular axe to grind, or that there are many distinguished psychologists who would disagree with the standards according to which this list has been compiled.

The creators of the Web site make the usual “EST” disclaimer that they are not passing judgment on therapies that are not on the list; but, perhaps because they are directing their arguments toward the general public, they do so in a more elaborate way, designed to appear “fair and balanced,” while clearly conveying the message that only the therapies on the list have been scientifically evaluated in controlled studies:

Some well-known psychotherapies do not appear here. Usually, this is because they have not been subjected to the types of controlled studies described above, rather than because these psychotherapies have been found to be ineffective or harmful. The field of psychotherapy research is a relatively new one, and therefore many therapies, which may prove to be beneficial, have simply not yet been studied. We recommend that consumers first seek out therapies that have been studied and shown to be beneficial in controlled studies. However, even the psychotherapies shown effective in controlled scientific studies do not help all patients. Therefore, if one of these treatments fails to help, it makes sense to try other therapeutic approaches, even when they have not been evaluated in controlled studies.

Does it require a randomized controlled trial to predict in which direction this ad-disguised-as-science will push most readers? Indeed, the Web site is quite explicit in directing readers first to “therapies that have been studied and shown to be beneficial in controlled studies” and clearly leaves the (false) impression that only the studies listed on the Web site meet that criterion.

More recently, “EST” advocates have begun an even more elaborate media blitz directed to the general public, with titles such as “Ignoring the Evidence: Why Do Psychologists Reject Science?” (Begley, 2009) and “Is Your Therapist a Little Behind the Times?” (Baker, McFall, & Shoham, 2009). These latter three authors, who seem to have provided much of the misinformation for Begley, a journalist, have also written an article in the publication *Psychological Science in the Public Interest*, this time directed toward a professional audience, in which the marketing mentality and self-interest that lies behind their declared championing of science does not require a trained clinician to detect. Early in the article they mention their concern that “psychologists are losing the opportunity to play a leadership role in mental and behavioral health care: Other types of practitioners are providing an increasing proportion of delivered treatment,” and state that “While the demand for mental health care is growing, psychologists are being bypassed as practitioners.” They do claim a page or two later that “The goal of reform would not be to secure employment for psychologists. Rather, it would be to increase the number of people who are helped by effective psychological interventions.” This might be a more credible claim if they did not also state, further on in the document, that “As a field, we have not done adequate market research,” or that “our view is that if an EST performs well relative to other *competitors for the health care dollar* . . . this finding retains public health and clinical significance” (italics added).

Part of their Walmart approach to mental health care is that “ESTs” are cheap because “many of these interventions can be disseminated without highly trained and expensive

personnel.”<sup>7</sup> Turning specifically to CBT, which is clearly the approach with which they are strongly identified, they state, as a virtue, that “CBT is effective even when delivered by nondoctoral therapists or by health educators with little or no prior experience with CBT who received only a modest level of training in that technique” (p. 38). In this they give short shrift to the large body of research attesting to the importance of the therapeutic relationship and to the skillfulness of the therapist, especially in the treatment of more difficult cases (e.g., Beutler et al., 2006; Gilbert & Leahy, 2007; Hofman & Weinberger, 2007; Norcross, 2002; Wampold, 2008). One must wonder if, in their own lives, they really would as readily entrust their troubled adolescent or their own struggles with relationship problems or with feelings of meaningfulness or satisfaction in life to a bachelors level therapist with a week’s experience.

### Beyond List Making and Horse Races: Toward a Focus on Principles and Processes

In response to the limitations of the “EST” approach, a range of writers widely respected for their commitment to rigorous research—a number of them associated with the cognitive–behavioral tradition—have suggested that we move beyond the “horse race” and “certifying” approach of the “EST” movement and focus instead on the *principles and processes* that underlie effective therapeutic work. Rosen and Davison (2003), writing in *Behavior Modification*, state the matter forcefully and pointedly: “any authoritative body representing the science of psychology should work toward the identification of ‘empirically supported *principles* of change’ (ESP); it should *not* list ‘empirically supported treatments’ (ESTs); and it certainly should never list proprietary, trademarked methods. . . . This situation makes for poor science and, at best, inefficient practice” (p. 303, italics added). Regarding training, they go on to add, “rather than students learning a list of cookbook methods that are tested and marketed for the treatment of a specific *DSM–IV* category, psychologists in training would learn basic principles of change and the range of their applications” (p. 306).

In a related vein, Allen, McHugh, and Barlow (2008), also writing from a cognitive–behavioral point of view, note that if training and research in psychological treatments continues to focus on discrete and separate treatment packages for each separate DSM diagnostic group, the task of training clinicians effectively to administer these treatments becomes daunting and impractical. In contrast, they argue, a unified, *principles-centered* approach addresses the very considerable comorbidity evident among emotional disorders. In support of their advocacy of a broad, principles-based approach (rather than what they call the “splitting approach” of the DSM, which is central to the “EST” movement), they review evidence indicating that even when treatments narrowly target single diagnostic categories, they tend to produce “significant improvement in additional comorbid anxiety and mood disorders that are not specifically addressed in treatment” (p. 220) Thus, it might be said, even the additional *successes* of “ESTs” challenge the premises on which the “EST” movement is grounded (especially its insistence on the medical model treat-

---

<sup>7</sup> I suppose here is the kernel of truth in their claim that their goal is not to “secure employment for psychologists.” Certainly they see little role for experienced or highly trained practitioners. In their scheme, therapists are low level technicians. The employment they seek to secure is for research psychologists like themselves.

ment-to-disease specificity and its refusal to consider findings deriving from more broadly targeted therapeutic efforts).

Other prominent researchers (e.g., Beutler, Clarkin, & Bongar, 2000; Castonguay & Beutler, 2003; Castonguay et al., 2006; Ehrenreich, Buzzella, & Barlow, 2007; Goldfried & Eubanks-Carter, 2004; Goldfried & Wolfe, 1996, 1998; Kazdin, 2002, 2007, 2008; Kazdin & Nock, 2003; Moses & Barlow, 2006; Pachankis, & Goldfried, 2007; Shapiro, 1995; Westen et al., 2004, 2005) have made similar or related points. Most of these researchers are strongly supportive of RCTs as *one element* in a broader overall research strategy to further the evolution of evidence-based clinical practice, but raise serious questions about the degree to which RCTs have dominated the attention of funding sources and the focus of the field. Beutler (2004), for example, states that the value of RCTs has “been overinflated by funding agencies who have favored this methodology in determining what studies will be funded” (p. 228). He goes on to note, in a fashion consonant with the present article’s emphasis on there being many more means of doing genuine science than are acknowledged in the generation of “EST” lists, that, “One need not dilute the scientific method to consider a wide array of treatment factors. One need only ensure that the method used is relevant to the question asked.” In this context, he describes a major research effort to “define explicit *principles* that will guide the clinician in the selection and conduct of treatment, rather than identifying *packages* and *manuals* of various treatment models.” Unlike the manualized packages that have been the focus of the “EST” movement, principles, he notes, “are more general and cross-cutting and thus can be applied to the integrated application of relationship qualities, participant factors, and treatment strategies” (pp. 228–229, italics added). By changing our focus in this way, he states, “we can circumvent the pervasive tendency to pit one treatment model against another, or to pit treatment model against relationship, in a dogma-eat-dogma competition” (p. 229).

### Overlapping Principles, Diverse Empirical Support

In thinking about the evidence base for a principles-based approach to psychotherapy, it is important to be clear that although the fundamental principles that guide therapeutic work from different theoretical orientations are by no means identical, there is far more overlap than is commonly assumed (Castonguay & Hill, 2007; Gold & Stricker, 2006; Hofman & Weinberger, 2007; Hubble et al., 1999; Norcross & Goldfried, 2005; Wachtel, 1997; Wampold, 2008). In part, these commonalities are often obscured by the widely diverging terminologies employed by therapists of each orientation. Consider, for example, Ehrenreich et al.’s (2007) cognitive-behavioral account of the basic principles they see as cutting across the treatment of a broad range of emotional disorders: (1) “as humans attempt to down-regulate or avoid their unexpected, excessive emotions they inevitably experience increasingly intense emotional states;” (2) “Emotionally Driven Behaviors (EDB) currently associated with a particular emotional experience must be resisted during exposure activities. Awareness of, and attention to, both the existing EDBs and the EDBs being developed are critical components of treatments for anxious and depressive disorders” (p. 200). Compare this account to Blagys and Hilsenroth’s (2000) depiction, based on an extensive review of the literature on psychotherapy process across different therapeutic approaches, of the distinctive features of psychodynamic-interpersonal approaches. The first two principles they list are (1) “a focus on affect and the expression of patients’ emotions;” and (2) “an exploration of patients’ attempts to avoid topics or engage

in activities that hinder the progress of therapy.” To be sure, the specific way that Barlow and his colleagues proceed is by no means identical to the ways that most psychodynamic therapists practice. The differences are obvious and noteworthy. But the shared principles of attending to the patient’s emotional experience and preventing *avoidance* of emotional experience (that is, in psychoanalytic terminology, of addressing *defenses*) point to some important commonalities of process behind the differences in form and procedure. In a related vein, the psychoanalytic concept of interpretation and the cognitive–behavioral concept of exposure derive from very different frames of reference and reflect different aims and visions of the therapeutic process, but a close examination of what is accomplished by effective interpretations indicates that there is considerable convergence in the psychological processes evoked in skillfully employing the two concepts (Wachtel, 1997, 2008).

Better understanding of these overlaps in the principles guiding therapeutic work from different vantage points is of value and importance for a number of reasons. From a clinical vantage point, therapeutic work rarely proceeds in smooth, uninterrupted fashion. Whether one employs the concept of resistance, or prefers some other terminology for the phenomenon, the task of helping people to confront the very experiences they have devoted so much of their lives to avoiding will almost inevitably yield moments when the therapist feels at sea, uncertain how to proceed. When continuing to do what one has been taught to do seems only to deepen the sense of impasse, it can be useful to appreciate that modes of interaction and intervention more characteristic of another orientation are not as alien as one has been taught. Discernment of the commonalities in principle and process beneath the surface differences in form can enable the therapist, whatever her orientation, to extricate herself from the tunnel vision that narrow identification with a particular orientation can create and thereby to find alternative means of interacting with the patient that accomplish similar ends. In thereby stretching or melding therapeutic techniques, new possibilities for effective therapeutic practice can emerge (Wachtel, 1993, 1997).

But appreciation of the commonalities in the principles that underlie therapeutic work from multiple orientations also enables us to have greater confidence in the validity of these principles. As discussed earlier in considering why overreliance on RCTs alone can be problematic, the findings of research are more credible and likely to endure if they derive from multiple perspectives and methodologies. When it becomes apparent that researchers who have operationalized their conceptualizations differently and approached their inquiry from different vantage points have obtained findings that point in a similar direction, we may have greater confidence in those findings and in the theoretical understanding that encompasses them.

Understanding of this overlap<sup>8</sup> between the principles and processes that underlie therapeutic work from different theoretical orientations may contribute as well to challenging a common misunderstanding—the idea that psychodynamic theory and practice are grounded solely in uncontrolled clinical observation and have no foundations in systematic controlled research. As Shedler (2010) has described,

---

<sup>8</sup> Again I wish to make clear that in pointing to unappreciated overlaps or commonalities I am not arguing that the principles underlying the various therapeutic orientations are *identical*. Indeed, there would be little point in attempting to forge an *integration* of the different approaches (Wachtel, 1997) if they were really all the same beneath the skin. There is more common ground than is usually appreciated, but each also brings different strengths, perspectives, and modes of intervention.

There is a belief in some quarters that psychodynamic concepts and treatments lack empirical support, or that scientific evidence shows that other forms of treatment are more effective. The belief appears to have taken on a life of its own. Academicians repeat it to one another, as do health care administrators, as do health care policymakers. With each repetition, its apparent credibility grows. At some point, there seems little need to question or revisit it because “everyone knows it to be so.”

With regard to the efficacy of psychodynamic therapy, Shedler demonstrates convincingly that what “everyone” knows to be so *is not* so—the evidential foundations of psychodynamic therapy are as strong as they are for any of the therapies on the “EST” lists, and, indeed, the evidence indicates that the psychodynamic therapies often achieve larger effect sizes and more enduring changes. But it is also important to be clear that much the same holds as well for the *basic principles* on which psychodynamic therapy is based. Indeed, the body of evidence in rigorous, controlled research that supports psychodynamic principles is so substantial that I cannot even cursorily review it here, but can only allude to a small fragment of the available literature.

Some of this research has been expressly conducted to address psychoanalytic concepts and/or reviewed for its explicit bearing on psychodynamic concepts and principles—see, for example, Cooper’s (1992) and Cramer’s (2008) reviews of research on defenses and the work of Andersen and her colleagues (e.g., Andersen & Chen, 2002; Andersen & Saribay, 2005; Andersen, Thorpe, & Kooij, 2007) on transference, as well as the more general reviews of Schut and Castonguay, 2001; Westen, 1998; Levy and Ablon, 2008; Blatt, 2008; Fonagy, Gergely, Jurist, and Target, 2005; Miller, Luborsky, Barber, and Docherty, 1993; Summers and Barber, 2010; and Curtis, 2009. Other large bodies of research bearing strongly on fundamental psychodynamic principles have been conducted by researchers with little or no interest in psychoanalysis per se. For many years, this work, conducted by researchers in fields such as social psychology, cognitive psychology, and cognitive neuroscience, assiduously avoided reference to unconscious processes. As Wilson (2002) put it, many academic psychologists “were reluctant to use the word ‘unconscious,’ out of fear their colleagues would think that they had gone soft in the head. Several other terms were invented to describe mental processes that occur outside of conscious awareness, such as ‘automatic,’ ‘implicit,’ ‘preattentive,’ and ‘procedural’” (p. 5). Now, however, he notes, “the terms ‘unconscious’ or ‘nonconscious’ . . . appear with increasing frequency in mainstream journals. . . . The gulf between research psychologists and psychoanalysts has thus narrowed considerably, as scientific psychology has turned its attention to the study of the unconscious” (p. 5). Although Wilson adds that, “This gap has not been bridged completely,” and that “the modern adaptive unconscious is not the same as the psychoanalytic one,” it is difficult to proceed in immersing oneself in this literature without recognizing that many of the fundamental assumptions that guide therapeutic work from a psychodynamic perspective are amply represented in this research (see, e.g., Banaji, Lemm, & Carpenter, 2001; Bargh, 2006; Petty, Fazio, & Brinol, 2008; Hassin, Uleman, & Bargh, 2005; Wilson, 2002).

### Conclusion: Science, Pseudoscience, and “ESTs”

There is an impressive body of evidence demonstrating the efficacy of a range of therapeutic approaches not on the “EST” lists. Much of that evidence includes RCTs. But the “EST” movement “disappears” them if they did not employ manuals or direct themselves to a narrowly defined DSM category. In a different kind of erasure, findings are “disappeared”

when “EST” advocates ignore the web of correlational studies, process-outcome studies, quasi-experimental studies, studies tracking the naturalistic occurrence of psychotherapy in the community with measurement at predetermined intervals, and so forth. Many of these naturalistic, quasi-experimental process-outcome studies are done with meticulous care and rigorously constructed controls for potentially confounding influences. When “EST” advocates treat these studies as irrelevant to the determination of what therapeutic approaches are “empirically supported,” they engage in a kind of deceptive casuistry similar to that which characterized for years the tobacco companies’ denial of the adverse health effects of cigarettes. There were, after all, no RCTs in which some people were required to smoke three packs a day and others were forbidden to place a single cigarette in their mouths. The evidence was overwhelming that cigarettes were harmful, but almost no single study, in any field of science, is utterly beyond questioning. By picking apart studies one by one with a relentless and tendentious tunnel vision, the overwhelming evidence presented by the converging indications of many different studies done from different angles with different methodologies was speciously refuted. Having their way with the evidence in similar fashion, “EST” advocates too play the role of disinterested scientists while presenting arguments that serve to obscure the preponderance of the evidence.

“EST” advocates have been much better at public relations than at science. Good science takes into account the full range of available evidence, probing constantly against smugly held certainties. Science that narrows its vision to exclude a wide range of studies that might challenge comfortably held beliefs is not science at all, but ideology dressed up in a lab coat for Halloween. Our profession, and the public, deserve better.

## References

- Ablon, J. S., & Jones, E. E. (1998). How expert clinicians’ prototypes of an ideal treatment correlate with outcome in psychodynamic and cognitive-behavioral therapy. *Psychotherapy Research, 8*, 71–83.
- Ablon, J. S., & Jones, E. E. (2002). Validity of controlled clinical trials of psychotherapy: Findings from the NIMH treatment of depression collaborative research program. *American Journal of Psychiatry, 159*, 775–783.
- Ablon, S. J., Levy, R., & Katzenstein, T. (2006). Beyond brand names of psychotherapy: Identifying empirically supported change processes. *Psychotherapy: Theory, Research, Practice, Training, 43*, 216–231.
- Allen, L. B., McHugh, R. K., & Barlow, D. H. (2008). Emotional disorders: A unified protocol. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders: A step-by-step treatment manual* (4th ed., pp. 216–249). New York: Guilford Press.
- Andersen, S. M., & Chen, S. (2002). The relational self: An interpersonal social-cognitive theory. *Psychological Review, 109*, 619–645.
- Andersen, S. M., & Saribay, S. A. (2005). The relational self and transference: Evoking motives, self-regulation, and emotions through activation of mental representations of significant others. In M. W. Baldwin (Ed.), *Interpersonal cognition* (pp. 1–32). New York: Guilford Press.
- Andersen, S. M., Thorpe, J. S., & Kooij, C. S. (2007). Character in context: The relational self and transference. In Y. Shoda, D. Cervone, & G. Downey (Eds.), *Persons in context: Building a science of the individual* (pp. 169–200). New York: Guilford Press.
- Baker, T., McFall, R., & Shoham, V. (2009, November 15). *Is your therapist a little behind the times?* Washington, DC: Washington Post.
- Banaji, M. R., Lemm, K. M., & Carpenter, S. J. (2001). The social unconscious. In A. Tesser & N. Schwartz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (pp. 134–158). Oxford, United Kingdom: Blackwell Publisher.

- Bargh, J. (2006). *Social psychology and the unconscious: The automaticity of higher mental processes*. New York: Psychology Press.
- Begley, S. (2009). Ignoring the evidence: Why do psychologists reject science? *Newsweek*, October, 12, 2009.
- Beutler, L. E. (2004). The empirically supported treatments movement: A scientist-practitioner's response. *Clinical Psychology: Science and Practice*, 11, 225–229.
- Beutler, L. E., Castonguay, L. G., & Follette, W. C. (2006). Therapeutic factors in dysphoric disorders. *Journal of Clinical Psychology*, 62, 639–647.
- Beutler, L. E., Clarkin, J. F., & Bongar, B. (2000). *Guidelines for the systematic treatment of the depressed patient*. New York: Oxford University Press.
- Blagys, M. D., & Hilsenroth, M. J. (2000). Distinctive activities of short-term psychodynamic-interpersonal psychotherapy: A review of the comparative psychotherapy process literature. *Clinical Psychology: Science and Practice*, 7, 167–188.
- Blatt, S. J. (2008). *Polarities of experience: Relatedness and self-definition in personality development, psychopathology, and the therapeutic process*. Washington, DC: American Psychological Association.
- Campbell, D. T., & Fisk, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Castonguay, L. G., & Beutler, L. E. (2006). Principles of therapeutic change: A task force on participants, relationships, and techniques factors. *Journal of Clinical Psychology*, 62(6), 631–638.
- Castonguay, L. G., & Beutler, L. E. (Eds.). (2003). *Empirically supported principles of therapeutic change*. New York: Oxford University Press.
- Castonguay, L. G., & Hill, C. (2007). *Insight in psychotherapy*. Washington, DC: American Psychological Association.
- Castonguay, L. G., Goldfried, M. R., Wiser, S., Raue, P. J., & Hayes, A. M. (1996). Predicting the effect of cognitive therapy for depression: A study of unique and common factors. *Journal of Consulting and Clinical Psychology*, 64, 497–504.
- Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, 52, 685–716.
- Chambless, D. L., Sanderson, W. C., Shoham, V., Bennett Johnson, S., Pope, K. S., Crits-Christoph, P., et al. (1996). An update on empirically validated therapies. *The Clinical Psychologist*, 49, 5–18.
- Cooper, S. H. (1992). The empirical study of defensive processes: A review. In J. W. Barron, M. N. Eagle, & D. L. Wolitzky (Eds.), *Interface of psychoanalysis and psychology* (pp. 327–346). Washington, DC: APA Books.
- Cramer, P. (2008). Seven pillars of defense mechanism theory. *Social and Personality Psychology Compass*, 2, 1963–1981.
- Curtis, R. C. (2009). *Desire, self, mind, and the psychotherapies: Unifying psychological science and psychoanalysis*. New York: Jason Aronson.
- Ehrenreich, J. T., Buzzella, B. A., & Barlow, D. H. (2007). General principles of the treatment of emotional disorders across the lifespan. In S. G. Hoffman & J. Weinberger (Eds.), *The art and science of psychotherapy* (pp. 191–210). New York: Routledge.
- Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2005). *Affect regulation, mentalization, and the development of self*. New York: Other Press.
- Gilbert, P., & Leahy, R. L. (Eds.). (2007). *The therapeutic relationship in the cognitive behavioral psychotherapies*. New York: Routledge.
- Goldfried, M. R., & Eubanks-Carter, C. (2004). On the need for a new psychotherapy research paradigm. *Psychological Bulletin*, 130, 669–673.
- Goldfried, M. R., & Wolfe, B. E. (1996). Psychotherapy practice and research: Repairing a strained relationship. *American Psychologist*, 51, 1007–1016.
- Hassin, R. R., Uleman, J. S., & Bargh, J. A. (Eds.). (2005). *The new unconscious: Social cognition and social neuroscience*. New York: Oxford University Press.

- Hofman, S. G., & Weinberger, J. L. (Eds.). (2007). *The art and science of psychotherapy*. New York: Routledge.
- Hubble, M. A., Duncan, B. L., & Miller, S. D. (Eds.). (1999). *The heart & soul of change: What works in therapy*. Washington, DC: American Psychological Association.
- Jones, E. E. (2000). *Therapeutic Action: A Guide to Psychoanalytic Therapy*. Northvale NJ: Jason Aronson.
- Jones, E. E., & Pulos, S. M. (1993). Comparing the process in psychodynamic and cognitive-behavioral therapies. *Journal of Consulting and Clinical Psychology, 61*, 306–316.
- Kazdin, A. E. (2002). The state of child and adolescent psychotherapy research. *Child and Adolescent Mental Health, 7*, 53–59.
- Kazdin, A. E. (2006). Arbitrary metrics: Implications for identifying evidence-based treatments. *American Psychologist, 61*, 42–49.
- Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology, 3*, 1–27.
- Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63*, 146–159.
- Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, 44*, 1116–1129.
- Kiesler, D. J. (1966). Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin, 65*, 100–136.
- Lamb, W., & Jones, E. (2009, January 19). *The effectiveness of psychoanalysis and longer-term psychodynamic psychotherapy: A meta-analysis*. Paper presented at the American Psychoanalytic Association meeting.
- Leichsenring, F. (2005). Are psychodynamic and psychoanalytic therapies effective?: A review of empirical data. *International Journal of Psychoanalysis, 86*, 841–868.
- Leichsenring, F., & Rabung, S. (2008). Analyzing effectiveness of long-term psychodynamic psychotherapy. *Journal of the American Medical Association, 300*, 1551–1565.
- Levy, R., & Ablon, S. J. (Eds.). (2008). *Handbook of evidence-based psychodynamic psychotherapy: Bridging the gap between science and practice*. New York: Humana Press.
- Luborsky, L., Diguier, L., Seligman, D. A., Rosenthal, R., Krause, E. D., . . . Schweizer, E. (1999). The researcher’s own therapy allegiances: A “wild card” in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6*, 95–106.
- Miller, N. E., Luborsky, L., Barber, J. P., & Docherty, J. P. (1993). *Psychodynamic treatment research: A handbook for clinical practice*. New York: Basic books.
- Moses, E. B., & Barlow, D. H. (2006). A new unified treatment approach for emotional disorders based on emotion science. *Current Directions in Psychological Science, 15*, 146–150.
- Norcross, J. (Ed.). (2002). *Psychotherapy relationships that work: Therapists contributions and responsiveness to patients*. New York: Oxford University Press.
- Norcross, J. C., & Goldfried, M. R. (Eds.). (2005). *Handbook of psychotherapy integration* (2nd ed.). New York: Oxford University Press.
- Pachankis, J. E., & Goldfried, M. R. (2007). An integrative, principle-based approach to psychotherapy. In S. G. Hoffman & J. Weinberger (Eds.), *The art and science of psychotherapy* (pp. 49–68). New York: Routledge.
- Petty, R. E., Fazio, R. H., & Brinol, P. (2008). *Attitudes: Insights from the new implicit measures*. New York: Psychology Press.
- Polanyi, M. (1958). *Personal knowledge: Towards a post-critical philosophy*. Chicago: University of Chicago Press.
- Polanyi, M. (1967). *The tacit dimension*. Chicago: University of Chicago Press.
- Rosen, G. R., & Davison, G. R. (2003). Psychology should list empirically supported principles of change (esps) and not credential trademarked therapies or other treatment packages. *Behavior Modification, 27*, 300–312.

- Roth, A., & Fonagy, P. (2004). *What works for whom? A critical review of psychotherapy research*. New York: Guilford Press.
- Sanderson, W. C. (2003). Why empirically supported psychological treatments are important. *Behavior Modification, 27*, 290–299.
- Schut, A. J., & Castonguay, L. G. (2001). Reviving Freud's vision of a psychoanalytic science: Implications for clinical training and education. *Psychotherapy: Theory, Research, Practice, Training, 38*, 40–49.
- Shapiro, D. A. (1995). Finding out how psychotherapies help people change. *Psychotherapy Research, 5*, 1–21.
- Shedler, J. (2010). The efficacy of psychodynamic therapy. *American Psychologist, 65*, 98–109.
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.
- Stricker, G., & Gold, J. (2006). *A casebook of psychotherapy integration*. Washington, DC: American Psychological Association.
- Sue, S., & Zane, N. (2006). In Norcross, J. C.; Beutler, L. E.; & Levant, R. F. (Eds.), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions*. (pp. 329–374). Washington, DC: American Psychological Association.
- Summers, R. F., & Barber, J. P. (2010). *Psychodynamic therapy: A guide to evidence-based practice*. New York: Guilford.
- Task Force on Promotion and Dissemination of Psychological. (1995). Procedures Training in and dissemination of empirically-validated psychological treatments: Report and recommendations. *Clinical Psychologist, 48*, 3–23.
- Wachtel, P. L. (1993). *Therapeutic communication: Principles and effective practice*. New York: Guilford Press.
- Wachtel, P. L. (1997). *Psychoanalysis, behavior therapy, and the relational world*. Washington, DC: American Psychological Association.
- Wachtel, P. L., & Schimek, J. G. (1970). An exploratory study of the effects of emotionally toned incidental stimuli. *Journal of Personality, 38*, 467–481.
- Wampold, B. E. (2008). *The great psychotherapy debate: Models, methods, and findings* (2nd ed.). New York: Routledge.
- Westen, D. (1998). The scientific legacy of Sigmund Freud: Toward a psychodynamically informed psychological science. *Psychological Bulletin, 124*, 333–371.
- Westen, D., & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology, 69*, 875–899.
- Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin, 130*, 631–663.
- Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2005). EBP ≠ EST: Reply to Crits-Christoph et al. (2005). and Weisz et al. *Psychological Bulletin, 131*, 427–433.
- Westen, D., Stirman, S. W., & DeRubeis, R. J. (2006). Are research patients and clinical trials representative of clinical practice? In Norcross, J. C.; Beutler, L. E.; & Levant, R. F. (Eds.), *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions*. (pp. 161–189). Washington, DC: American Psychological Association.
- Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA: Harvard University Press.