

# What Is the Scientific Meaning of Empirically Supported Therapy?

T. D. Borkovec and Louis G. Castonguay  
Pennsylvania State University

It is important to define precisely what is and is not meant by "empirically supported treatments," rigorously based on what is actually known about the nature of experimental therapy research. The criteria for empirically supported treatments merely allow conclusions about whether treatments cause any change beyond the causative effect of such factors as placebo or the passage of time. Applied implications are limited, due to external validity and to the fact that applied decisions are influenced by cost-benefit analyses. Creating increasingly effective therapies through between-group designs is best done by controlled trials specifically aimed at basic questions about the nature of psychological problems and the nature of therapeutic change mechanisms. Naturalistic research is important for external validity but is valuable only if it uses scientifically valid methods to address basic knowledge questions.

We love scientific research. There is a type of precision and beauty that is not present in other ways of acquiring knowledge. Empirical relationships discovered from carefully conducted experimental studies stand as relative truths that show us how things are interrelated at a concrete level and how theories are in need of revision at the conceptual level. In both ways, they provide us with opportunities to perceive and behave in our worlds in increasingly accurate and adaptive ways.

When we engage in the scientific enterprise, we agree to follow the same rules of evidence so that we or anyone else who knows these rules can agree on the knowledge so obtained. Of course, there remain numerous areas of potential debate in interpreting the results of any experimental investigation, how they comment on theory revision, and in what ways their demonstrated relationships apply to things beyond the specific questions and circumstances of the investigation. But because we agree in general on issues of measurement, design, methodology, statistics, and how these features affect what we can and cannot conclude, sufficient accord exists to allow confident and continuous progression of increasingly accurate knowledge from which increasingly useful applications can emerge. The history of psychological research (both basic research and therapy outcome research) on which the Task Force on Promotion and Dissemination of Psychological Procedures (1995) report and its updating article (Chambless et al., 1996) owe their foundation is a testimony to this.

When the results of scientific studies are applied to new and important questions that may directly or indirectly affect clinical training, clinical treatment, and financial decisions about how to treat, it is useful for us to return to our roots in empirical

science and to carefully consider again the nature of our scientific methods and what they do and do not provide in the way of possible conclusions relevant to those questions. We face such a moment now in addressing the ways in which controlled therapy trials provide information relevant to important questions concerning the effectiveness of psychotherapy.

## The Nature of Experimental Research

Let's first consider what experimental therapy outcome research is capable of doing. Like any form of experimental research, a well designed and conducted therapy outcome study allows us merely to demonstrate a cause-and-effect relationship. Competing causes are ruled out by holding everything constant among comparison conditions except the manipulated variable. Potential causes common among comparison conditions cannot explain observed differences in outcome; potential causes present in one condition and not in the other remain unrejected and thus represent both the likely source of causal influence on the outcome differences and the promising site for further experimental pursuit of even more specific cause-and-effect relationships.

We can, for example, randomly assign clients to therapy and no-treatment conditions, thereby equating two groups of people on all factors except for the presence or absence of therapy. If the two conditions differ in outcome, we can reject several competing factors (history, maturation, repeated testing, instrument drift, statistical regression, attrition, selection bias, and interactions between selection bias and other factors; cf. Campbell & Stanley, 1963) as the causal explanation of not the amount of change in the therapy group but of the difference in degree of change between the two conditions. There was something contained in "therapy" and not in no-treatment that was "effective" (i.e., caused additional change) beyond such factors as the mere passage of time or the effects of repeated testing. From a scientific perspective, such a result would suggest that further experimental pursuit of that "something" will be worthwhile.

Despite the fact that several competing factors could be ruled

---

T. D. Borkovec and Louis G. Castonguay, Department of Psychology, Pennsylvania State University.

Preparation of this article was supported in part by National Institute of Mental Health Research Grant MH-39172.

Correspondence concerning this article should be addressed to T. D. Borkovec, Department of Psychology, Pennsylvania State University, University Park, Pennsylvania 16802. Electronic mail may be sent to [tdb@psu.edu](mailto:tdb@psu.edu).

out by this design, the therapy and no-treatment conditions differed in several further ways, and any one or a combination of these ways may thus be the causal factor in the therapy's superior outcome. So we can draw a cause-and-effect conclusion from such a design, but it is very limited in its ability to specify the causal factors involved. The experimental solution is to create another comparison condition that contains more elements (more possible causative factors) in common with the therapy condition so that the groups differ in fewer ways. Then any observed difference in outcome would not be due to these factors but rather to what was specific and unique to the therapy. The use of placebo conditions has historically been an important example (but often misused and frequently inappropriate for several reasons; cf. O'Leary & Borkovec, 1978) of an attempt to hold constant such further factors as client expectancy for improvement, contact with a therapist, suggestion and demand characteristic effects, and relationship with a caring person.<sup>1</sup> Comparisons to alternate therapies can serve this same purpose, as long as additional methodological requirements are met (e.g., equivalence of conditions in initial credibility and expectancy for improvement; cf. Borkovec, 1994). Differences in outcome between therapy and placebo (or alternate therapy) groups would allow us to rule out factors common to all therapeutic relationships as the causal explanation of the greater degree of change found for the therapy condition and to conclude that something specific contained within the therapy and beyond the effects of these common factors caused such change. We can say that therapy was more "effective" (caused a portion of additional change) than the provision of a general therapeutic relationship and whatever (as yet unspecified) causative factors such a relationship might possibly contain.

Given the above discussion, we can identify the central underlying assumptions of the earlier Task Force criteria as well as the newly suggested criteria offered by Chambless and Hollon (1998) and then evaluate more precisely what those criteria yield in terms of possible conclusions. The type of experimental design and methodology indicated in the Task Force report for empirically supporting the efficacy of a therapy (or for supporting a therapy as "efficacious and specific" by Chambless and Hollon's new criteria) would allow us to draw one and only one unambiguous conclusion: In a criterial study, therapy caused a degree of change beyond the amount of change caused by factors common to all therapies (and, of course, beyond chance). The results might contain much more information than that, but we are not in a position to be unambiguously certain about the validity of such inferences. We can only conclude that (a) the therapy contained some (as yet unknown) active ingredients that actually caused some degree of change (specifically, that degree of change beyond the change caused by factors common to all psychotherapy or placebo conditions or chance) and (b) this was true for the particular clinical problem, clients, setting, methods, therapists, and ways of measuring improvement that were used in the investigation. For scientific purposes, this would be a very important conclusion. Rigorous scientific method indicating fairly unambiguously that the therapy contained specific causes of improvement gives promise that significant basic knowledge about the nature of the psychological problem being treated and the mechanisms of change by which it is improved can be gained. But superiority of a therapy

over common therapy factors is only the starting point for useful scientific research of this type. Such a result encourages further experimental investigation of the therapy by means of powerful experimental designs that evaluate the causative contributions of the separate and combined elements of a therapy (dismantling or component control designs), of the addition of a new therapy element to an already established therapy (constructive or additive designs), and of levels of dimensions of therapeutic process thought theoretically to mediate a technique's causative effects (parametric designs). These designs allow increasingly specific cause-and-effect conclusions and thus markedly enhance our basic knowledge ("this causes that") about therapeutic change. What further applied implications can be drawn beyond the one scientifically allowable conclusion about cause and effect possible from investigations that meet the empirically supported treatment criteria is a quite separate question.

The same comments are applicable to Chambless and Hollon's (1998) new category of "established in efficacy" or "efficacious" (which basically refers to treatment/no-treatment designs), except that the possible conclusions from its criterial studies are even more severely limited in their basic knowledge and applied implications. Studies using this type of design would allow us only to conclude that the therapy caused a degree of change beyond the amount of change caused by such factors as the mere passage of time or the effects of repeated testing.

#### The Relevance of Controlled Therapy Outcome Research to Questions of Applied Efficacy

Given that the nature, purpose, and capability of experimental therapy outcome research involve the demonstration of increasingly specific cause-and-effect relationships to acquire basic knowledge, how do such controlled trials relate to the extremely important applied question of "efficacy" or "effectiveness" as that question is asked by our profession (e.g., for training purposes), consumers, third party payers, and society in general? That question comes in one of three forms: (a) Is a therapy effective? (b) How effective is a therapy? and (c) Which therapy is most effective? It is useful to carefully specify precisely what experimental research can say about each of these ways of asking these applied questions and to distinguish this clearly from its above-stated scientific goal.

#### Is a Therapy Effective?

Historically, and for the most part currently, this form of the applied question is often asking whether a therapy is better than

<sup>1</sup> The phrase, *nonspecific factors*, has been traditionally used to describe the processes presumed to operate commonly in both placebo conditions and in most forms of psychotherapy. Several authors, however, have argued that this phrase is a misnomer that should be eliminated (e.g., Castonguay, 1993). Accordingly, we use the phrase, *common factors*, in this article to refer to this set of factors that are causatively related to improvement occurring in placebo conditions and, therefore, to some of the therapeutic improvement observed in therapy investigations. It should be kept in mind, however, that there are also factors common to many psychotherapies that go beyond those factors contained within placebo conditions (e.g., treatment strategies to facilitate client awareness, corrective experiences, or continuous reality testing).

nothing at all (answered experimentally by the no-treatment design), although more well-informed questioners will increasingly demand better evidence than this. From the earlier discussion of each type of design, the best we can say is that controlled trials have the capability of empirically demonstrating whether or not a therapy does indeed contain some causative factors (beyond certain other factors) that lead to clinical change. This can be an important answer to this form of the applied question. However, the significance of that answer partly depends on what types of comparison conditions containing those "certain other factors" have been used. Therapies that do not cause change beyond the passage of time are likely not worth any time or financial investment for either clinical application or research pursuit. Therapies that cause change beyond the passage of time but do not cause change beyond the provision of common factors may be clinically worthwhile, but no more so than any other such treatment containing those factors. And their scientific value likely resides solely in their potential use in future experimental investigations deliberately aimed at specifying the causative factors responsible for placebo effects. If a therapy is superior to common factors as represented by placebo or alternate therapy conditions, then it is scientifically very useful for further experimental pursuit to isolate its specific causative ingredients, and it may be clinically useful. Therapy elements demonstrated to play no causative role in dismantling, parametric, and additive designs likely have little or no value for either clinical or scientific purposes, whereas those elements found to be causative may be clinically worthwhile and are scientifically very important because of the degree of specificity of the causal factors thereby identified.

Assume, for example, that we conduct a dismantling design on systematic desensitization for phobias (relaxation alone vs. graduated imaginal exposure alone vs. graduated imaginal exposure during a relaxed state) and find equivalence between the two exposure conditions and superiority of both to relaxation alone. The applied implication (assuming generalizability of the results to the applied setting) is that we need not waste client and therapist time and money on 10 sessions of relaxation training when treating phobias. The basic science implication is that experimental pursuit of the mechanisms of change in phobias by means of exposure methods would not usefully include relaxation in its future theoretical or empirical quests. On the other hand, if relaxation is demonstrated to play a causative role (e.g., graduated imaginal exposure during a relaxed state is found to be superior to either element alone), then it may be useful in clinical applications. Scientifically, such a demonstration would encourage further experimental pursuit of why relaxation contributes to outcome, ideally within some relevant theoretical context. For example, its facilitative effects may relate to attentional processes, psychophysiological processes, depth of emotional processing, or some other internal condition that modulates the influence of Pavlovian extinction procedures in humans. In fact, a variety of threads of research findings now indicate that, under some circumstances, (a) relaxation does facilitate phobic extinction process, (b) relaxation training increases parasympathetic tone, (c) parasympathetic tone relates to attentional deployment, and (d) cognitive and emotional processing is facilitated by the presence of a relaxed state. Notice that we are acquiring basic knowledge about human behavior and experi-

ence that is potentially broader in its implications than merely Pavlovian theory and behavioral extinction applications. Moreover, such results would suggest additional basic and applied possibilities for future experimental (cause-and-effect) pursuit. For example, the presence of a relaxed state may well facilitate emotional processing in other therapies grounded in other theoretical accounts of human change process (e.g., the accessing of previously denied or suppressed emotional experience in experiential therapy).

### *How Effective Is a Therapy?*

In each of the above design examples, we stated that differential outcomes between the involved comparison conditions may suggest the clinical utility of the superior condition. This is because experimental trials can answer directly questions about whether a therapy condition is effective and how effective it is only in a limited way by stating whether that condition causes change beyond the degree of change caused by whatever factors are contained in the comparison group. Applied implications relating to how effective the superior condition is are indirect products from such results, and those implications must be evaluated in the context of at least three further considerations. First, finding a statistically significant difference does not tell us how large that difference is. For meaningful applied implications, we need also to evaluate such things as effect size or difference in degree of clinically significant change (Jacobson & Traux, 1991; Kendall, 1997). Task Force criteria are silent on this issue, allowing for the possibility that a therapy could be labeled *empirically supported* even though, for example, it does not yield a greater percentage of clients reaching high endstate functioning than the comparison condition. Second, how much difference a therapy condition causes beyond the amount of change caused by comparison groups must be evaluated within the context of temporal or financial cost-benefit analyses or both. An element of therapy (e.g., presence of relaxation during phobic stimulus exposures) may be demonstrated by controlled trials to cause some of the change, but the amount of difference may not be large enough to justify spending 10 hours of therapy time to train the client in deep relaxation methods. Exactly what "large enough" is (relative to its cost) has not been systematically addressed in either a statistical or a clinical utility sense, even though this will surely be of major importance in the future. Third, even if a therapy is demonstrated to cause a greater degree of change (by whatever measurement definition) than a comparison condition, we remain uncertain about the external validity of this result, whether we are speaking of generalizability of results in general or of tests of the transportability of specific protocol manuals. In typical efficacy studies, restrictions routinely exist regarding such features as inclusion-exclusion criteria for client admission, choice of protocol therapists, and supervision of therapists with rigorous adherence controls. Because other authors in this issue will undoubtedly comment in detail on how this critical issue might severely limit the relevance of many controlled therapy trials meeting empirically supported treatment criteria, we do not discuss it further.

### *Which Therapy Is Most Effective?*

Of course, the Task Force (1995) and the Chambless and Hollon (1998) article did not intend to address this question, and

nothing about their reports relates explicitly to it. Their articles do relate to this question in an indirect and important way, however. If a therapy is superior to another treatment or matches the outcome of an already established treatment with adequate statistical power, then that therapy is said to have adequate empirical support. Thus, comparative designs (the comparison of one therapy technique to a very different therapy technique, e.g., interpersonal psychotherapy and cognitive behavioral therapy for depression) represent one vehicle for demonstrating empirical support for a therapy by the criteria. Within the Task Force's context, this methodological reasoning is sound, as long as we remember that the precise meaning of the outcomes from such designs resides solely in the demonstration that a therapy causes some change beyond that caused by placebo effects or chance. What is troubling is that the criterial use of comparative designs for this valid reason and purpose may reinforce the idea that comparative designs are useful methodological devices in therapy research for other purposes. Indeed, comparative designs are never conducted merely to control for placebo effects; they are explicitly trying to answer the question, "Which therapy is more effective?" Unfortunately, these designs are not internally valid (therefore they cannot answer this or any other question beyond the ruling out of placebo effects), and even if they were valid, their results would have inadequately brief applied relevance. Although the arguments in defense of these statements are lengthy (cf. Borkovec, 1994; Borkovec & Miranda, 1996), they can be summarized briefly: (a) Comparative designs do not even approximate the experimental ideal of holding all but one factor constant among compared conditions. The two compared therapies differ from one another in a very large number of ways. No rival hypotheses about what caused any outcome difference (other than common factors) can be ruled out, and we thus acquire very little scientific knowledge. (b) One of the ways in which the two therapies may differ is absolutely crucial to the internal validity of the comparison: They must be provided with an equivalent degree of quality; if not, the comparison is confounded. As yet, we do not have valid and reliable measures of quality for any therapy, much less for two or more compared therapies. Employing expert therapists from each therapy tradition does not solve the problem because this inherently confounds the design with therapist characteristics, and the absence of quality measures still precludes demonstration of equivalent quality. (c) Therapy outcome trials require 3–5 years to complete. By the time results are obtained, each therapy has changed on the basis of new clinical and empirical knowledge. So the results of such designs, even if they were internally valid, would be relevant only to the ways in which those therapies were operationally defined by manuals 3–5 years earlier. These criticisms are not equally applicable to dismantling, additive, and parametric designs. In the latter studies, many more elements can be held constant, and therapists are equally trained in all conditions and see equal numbers of clients in each condition. Moreover, with a basic knowledge focus and goal, established empirical relationships and the laws of behavior on which they comment are specifiable and not short-lived in significance.

There is a final troubling aspect to comparative designs that retards progress in acquiring knowledge about therapy. They encourage conflict and debate that have more to do with political agendas and personal theoretical beliefs than with knowledge

acquisition. It is amazing how much time and energy has been devoted to discussions about which therapy is more effective on the basis of investigations using inherently invalid comparative designs. Although comparative studies might provide a useful controlled context for the systematic investigation of causative process factors common and unique to different therapies, we fear that the natural but erroneous inclination to draw some significance from any differential outcome will persist.

### Recommendations for Future Therapy Outcome Research

Our concern about the empirically supported therapy movement is (a) that we may draw erroneous conclusions from outcome results when applying them to important applied questions and (b) that our focus in designing and conducting therapy outcome studies in an effort to answer these questions draws our attention away from how controlled therapy trials can best contribute to the development of more effective therapies (the very goal that we are seeking to reach) and thereby affects how we actually design and conduct therapy research. With regard to the first concern, we have argued that therapy research, as experimental science, has the capability of establishing cause-and-effect relationships and nothing more. Statements about empirical support for a therapy technique on the basis of studies meeting empirically supported treatment criteria and any implications drawn from such statements must be clearly made with reference to the only defensible conclusion: The investigated therapy causes some improvement beyond chance and factors common to all therapeutic relationships (for efficacious and specific therapies) or beyond such factors as the passage of time and repeated testing (for therapies established in efficacy). This is the only meaning of *empirically supported* within the context of the empirically supported treatment criteria. With regard to the second concern, we would like to suggest in the remainder of this article how we might usefully conceive of and implement between-group experimental designs, such that they can most rapidly and significantly generate basic knowledge on which to base new therapeutic developments and, as a consequence, indirectly contribute to answering the applied questions that are so important to our profession. These suggestions fall into two domains: the need for programmatic therapy outcome investigations deliberately designed to acquire basic knowledge and the need for therapy research in naturalistic settings whose primary goal is also to answer basic questions.

#### *Therapy Research as Basic Science*

Controlled therapy trials are potentially powerful empirical vehicles for the acquisition of basic knowledge. This is not to say that controlled trials do not have the capability of contributing to the development of increasingly effective therapies. Indeed, this is precisely one of their ultimate and most important goals. But the achievement of that goal is based on a proper understanding and application of the goals and methods of experimental science. The identification of increasingly specific cause-and-effect relationships leads to better theoretical understandings of the nature of the psychological problems being treated and the nature of the mechanisms of change underlying any demonstrated

causative roles for a therapy, its elements, its parameters, or elements added to it. From this knowledge, hypotheses about modifications or additions to a therapy emerge and can be tested. In this programmatic way, we will learn about additional causes of change and which combinations of therapy elements can yield the greatest amount of change because they contain more therapeutic causes. The approach most likely to pay off in achieving such ends recommends that we go deeply into what we know. At the theoretical level, this means adopting the strong inference perspective on the pursuit of basic knowledge: Generate rival hypotheses, design experiments that will allow the ruling out of one or more of those rival hypotheses, conduct a clean experiment, and re-cycle these steps on hypotheses that remain unrejected (Platt, 1964). At the concrete level of between-group designs, this means making use of dismantling, constructive, and parametric designs whose very process leads to the identification of increasingly specific cause-and-effect relationships for theoretical and applied purposes. In addition to accomplishing these goals, the beauty of these designs includes the facts that they control for common factors (thus eliminating the need for questionable placebo conditions) and they significantly lessen the likelihood of potential confounds from differential client expectancy or therapist quality (cf. Borkovec, 1994).

Such design approaches are, moreover, not theory-specific. Although they have been most often used with cognitive behavioral therapies, they are just as applicable to psychodynamic and experiential therapies whose own theoretical underpinnings suggest parameters, processes, and techniques that can be varied or added to existing methods to identify causative contributions. It was once argued that protocol manuals could not in principle be developed for these therapies; this turned out not to be the case. In the same way, the fact that these designs have not often been used in the past for some types of therapy does not mean that adaptations of the designs cannot be constructed and applied to them as a between-group means of acquiring basic knowledge. Using a dismantling design, for example, one could determine whether expressive techniques (e.g., interpretation, confrontation) have a causal effect on a client's improvement in psychodynamic treatment, above and beyond supportive interventions (e.g., education, intellectual guidance), by comparing two treatment protocols: a psychodynamic therapy with expressive and supportive elements and a psychodynamic therapy containing only the supportive interventions. A parametric design comparing two gestalt therapy conditions varying in terms of the degree of focus on, or resolution of, unfinished business would provide a rigorous test of a theoretically assumed process of change. Adding an experiential technique such as evocation of feelings to a psychodynamic treatment and comparing this integrative therapy protocol to a traditional psychodynamic therapy could not only assess the causal impact of the added techniques but also evaluate the theoretically assumed role of emotional deepening as a mechanism of change in psychodynamic treatment.

We have emphasized how between-group therapy designs can be best interpreted and can best contribute to the evolution of increasingly effective therapies, because that was largely the context of Task Force criteria and the listing of example therapies meeting those criteria. It is also important to mention that basic knowledge and applications will also derive from correla-

tional research methodologies that investigate predictor, mediator, and moderator variables related to outcome and that pursue the identification of causative factors through process research. These approaches are likely to play a very significant role, in addition to between-group designs, in future (and much needed) therapy research in naturalistic settings.

### *Therapy Research in the Naturalistic Setting*

Irrespective of abstract debates surrounding the question of whether controlled laboratory outcome investigations are generalizable to the practice setting, the only way to find out whether the results of between-group investigations do generalize to the applied clinical setting is to test generalizations empirically in that setting. The customary nature of applied research is to evaluate applications in specific environments to solve applied problems specific to those environments. For applied therapy work, this would mean the testing of therapies, developing specialized applications of therapies, and evaluating those applications with the particular therapists, clients, and psychological problems characteristic of the particular agency. The ultimate answers to the kinds of questions that the empirically supported treatment movement is trying to address must be found here. But we must again carefully consider what the meaning of empirical results obtained in such settings would be. Except for the fact that naturalistic setting research can reduce concerns about generalizability, the same bottom line emerges that we discovered when discussing controlled outcome and process research, that is, the need for a basic science approach.

Let us first assume an idyllic world where valid and reliable assessments of outcome are obtained at pretherapy, posttherapy, and follow-up periods for every client treated by psychological methods in the entire country. What evidence would we have that therapy is effective? We could get a concrete number (e.g., 60% of clients are returned to normal functioning after therapy), but what would this number mean? Does it represent powerful, moderate, or poor effectiveness? We would of course want to break the data down into such categories as type of presenting problem and type of therapy administered to be more specific about improvement rates, but we are still left with the same interpretive difficulty. Numbers are meaningful only in relative comparison. Would that percentage of improvement have occurred merely with the passage of time or with a friend who is a sympathetic listener? As soon as we raise these kinds of questions, we immediately return to the need for the scientific methodology that was created specifically to rule out such rival interpretations of what was causing change. The implication is this: It is essential that our profession begin a serious effort to conduct scientifically valid therapy research in the applied setting, both to address the generalizability issue and to generate meaningful results of long-lasting value. And "meaningful results" means findings that systematically contribute to the discovery of specific cause-and-effect relationships that directly yield basic knowledge and thus lead indirectly (but powerfully) to the development of increasingly effective (causative) therapies.

A demonstration project of this type is currently under way. The Pennsylvania Psychological Association has established a Practice Research Network to organize practicing clinicians and clinical researchers for the sake of conducting therapy

research in the practice setting. Thus far, over 200 clinicians from around the state have volunteered to participate. With pilot funds provided by the American Psychological Association, the Network has put together a core assessment battery that will be given to all clients seen by participating clinicians at pretherapy, midtherapy, posttherapy, and 6-month follow-up. This, in and of itself, is insufficient, as discussed above. Once such an infrastructure is place, however, specific and scientifically meaningful basic research projects can be implemented. Here, the possibilities are legion. The core battery already assesses numerous client demographic and intra- and interpersonal psychological variables at pretherapy as well as several therapist characteristics (e.g., years of experience and the degree to which various theoretical orientations contribute to case conceptualization and treatment of clients). Other assessment devices will be added later in response to theory-driven and clinician-driven questions as those emerge. Correlations between these variables and immediate and long-term outcomes will provide initial guidance in the pursuit of cause-and-effect relationships, by ruling out some relationships and encouraging the experimental evaluation of others. Process research in conjunction with the assessment battery has a potentially enormous contribution to make in isolating likely mechanisms of change associated with client experiences and behaviors, therapist moment-to-moment interventions, and the therapeutic relationship itself. The coding from session audiotapes of such processes as depth and type of client affective experience (e.g., by means of the Experiencing Scale; Klein, Mathieu, Gendlin, & Kiesler, 1969), therapist focus on intrapersonal versus interpersonal themes (e.g., by means of the Coding System of Therapeutic Focus; Goldfried, Newman, & Hayes, 1989), client-therapist interactions (e.g., by means of the Structural Analysis of Social Behavior; Benjamin, 1974), and specification of the actual techniques used by the clinician (e.g., by means of adherence checklists previously developed for several therapy orientations in controlled trials) can provide rich information from many client-therapist dyads in a form compatible with how many clinicians customarily view their work with clients. The rigorous integration of qualitative and quantitative methods (e.g., task analysis) can provide theoretically driven, contextual, and sequential analyses of the process of change taking place (or failing to take place) in success and failure cases identified by means of the core assessment battery. Although labor-intensive and time-consuming, process research can provide support for mechanisms of change assumed to be operating in different approaches and can generate further hypotheses about causal factors of change. Furthermore, as cogently argued by Grawe (1997), the use of process findings to guide modifications of current psychotherapies may represent the best strategy to increase the effectiveness of such treatments.

The core of specific research studies within this infrastructure, however, will involve true randomized trials that allow specific cause-and-effect demonstrations in scientifically valid ways, on the one hand, but are also sensitive to the clinical, ethical, and pragmatic realities of the practice setting, on the other. For example, once a sufficient baseline period of assessments has been accumulated, requests for volunteers for a specific, proposed study will be sent to all clinicians participating

in the core assessment battery project. Those who volunteer will be randomly assigned to a training condition and a delayed-training condition. The former clinicians will receive workshop training and phone supervision in an experimental intervention method, and all of the clients in relevant diagnostic groups will then be treated by the clinicians' usual methods, plus the addition of the experimental element. Delayed-training clinicians will continue to treat their clients by their usual methods. After a year of data collection, the delayed-training group will receive the training and then add that element to their treatment of subsequent clients for the next year. Thus, an additive design, with a cross-over replication of the effects of the experimental intervention, can be accomplished. The actual content in such trials can focus on any number of questions. For example, is protocol therapy by manuals developed in previous controlled outcome investigations transportable to the practice setting? Can the presence of a deeply relaxed state impact on the efficacy of experiential therapies or of cognitive therapies? Can initial cognitive therapy facilitate the depth of emotional processing during experiential sessions, and vice versa? Does increased therapist focus on interpersonal themes or on childhood themes increment the effects of treatment-as-usual? Would specialized training in working with hostile and controlling clients increase treatment efficacy or decrease early termination? Not only would such controlled outcome studies answer important basic questions relevant to real clinical concerns, but they would also provide the opportunity to integrate outcome research with the kinds of process research described above. And in so doing, the opportunity is created for (a) the testing of various therapy elements for both common and unique mechanisms of change, as described by the theories for several therapeutic orientations, and (b) empirically driven integrations of psychotherapy traditions and their techniques.

From our perspective, therapy research is not a matter of sequenced investigation, beginning in the laboratory and then moving to the field, or vice versa. Therapy research should be conducted in both settings with the same goal, the acquisition of basic knowledge. Each setting has its relative advantage (greater control in the laboratory and thus potentially greater specificity of conclusions, greater generalizability in the naturalistic setting to other applied settings).

### *Concluding Comments*

We really do believe that this is a route to knowledge that can yield new and better therapies. We also believe that our field has largely failed to emphasize the scientific perspective sufficiently in this way (either in controlled efficacy studies or in naturalistic effectiveness studies) to accomplish this goal. Part of the reason for this failure is that we quite naturally wish to see therapy studies directly answering the profoundly important applied questions, and so we design and conduct such studies with this goal in mind rather than with the goal of acquiring basic knowledge. Demands from groups external to our profession and internal demands on ourselves to demonstrate the value of what we as practitioners do and to develop better ways of doing it, however, place the emphasis in exactly the wrong direction for accomplishing this very goal.

We support letting society and our profession know what

therapies we have that are indeed effective, and controlled experimental studies listed by the Task Force reports represent a source of evidence for this. But let us be clear to society and our profession both about what we mean (more causative than placebo or no-treatment factors) and about what we cannot conclude. And let us get on with the business of developing and evaluating increasingly effective therapies. Our fear is that misunderstandings about, or misapplications of, the real scientific meaning of studies meeting empirically supported treatment criteria, the internal and external forces that led to the movement's initial work, and reactions from within and outside of the profession, all take our attention away from what we can and should be doing as empirical scientists and away from the deployment of the most powerful experimental methods that we have that can lead to answers to the very questions that are being raised.

### References

- Benjamin, L. S. (1974). Structural analysis of social behavior. *Psychological Review*, 81, 392-425.
- Borkovec, T. D. (1994). Between-group therapy outcome design: Design and methodology. In L. S. Onken & J. D. Blaine (Eds.), *NIDA Research Monograph No. 137* (pp. 249-289). Rockville, MD: National Institute on Drug Abuse.
- Borkovec, T. D., & Miranda, J. (1996). Between-group psychotherapy outcome research and basic science. *NIMH Psychotherapy and Rehabilitation Research Bulletin*, 5, 14-20.
- Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally.
- Castonguay, L. G. (1993). "Common factors" and "nonspecific variables": Clarification of the two concepts and recommendations for research. *Journal of Psychotherapy Integration*, 3, 267-286.
- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66, 7-18.
- Chambless, D. L., Sanderson, W. C., Shoham, V., Johnson, S. B., Pope, K. S., Crits-Christoph, P., Baker, M., Johnson, B., Woody, S. R., Sue, S., Beutler, L., Williams, D. A., & McCurry, S. (1996). An update on empirically validated therapies. *Clinical Psychologist*, 49, 5-18.
- Goldfried, M. R., Newman, C. N., & Hayes, A. M. (1989). *The coding system of therapeutic focus*. Unpublished manual, State University of New York at Stony Brook.
- Grawe, K. (1997). Research-informed psychotherapy. *Psychotherapy Research*, 7, 1-19.
- Jacobson, N. S., & Traux, P. (1991). Clinical significance approach to defining meaningful change in psychotherapy. *Journal of Consulting and Clinical Psychology*, 59, 12-19.
- Kendall, P. C. (1997). Editorial. *Journal of Consulting and Clinical Psychology*, 65, 3-5.
- Klein, M. H., Mathieu, P. L., Gendlin, E. T., & Kiesler, D. J. (1969). *The Experiencing Scale: A research and training manual* (Vol. 1). Madison: University of Wisconsin Extension Bureau of Audiovisual Instruction.
- O'Leary, D. K., & Borkovec, T. D. (1978). Conceptual, ethical, and methodological problems of placebo groups in psychotherapy research. *American Psychologist*, 33, 821-830.
- Platt, J. R. (1964). Strong inference. *Science*, 146, 347-353.
- Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically-validated psychological treatments. *Clinical Psychologist*, 48, 3-23.

Received September 30, 1996

Revision received January 21, 1997

Accepted February 18, 1997 ■