

# Stressing the (Other) Three Rs in the Search for Empirically Supported Treatments: Review Procedures, Research Quality, Relevance to Practice and the Public Interest

John R. Weisz and Kristin M. Hawley, University of California, Los Angeles

Paul A. Pilkonis, University of Pittsburgh School of Medicine

Sheila R. Woody, Yale University

William C. Follette, University of Nevada, Reno

**The Society of Clinical Psychology's task forces on psychological intervention developed criteria for evaluating clinical trials, applied those criteria, and generated lists of empirically supported treatments. Building on this strong base, the task force successor, the Committee on Science and Practice, now pursues a three-part agenda: (a) evolution of review and classification procedures with an emphasis on reliability across reviewers, (b) an active role as gadfly in promoting improved research, and (c) a dissemination program (with an evolving web site) to make our process, findings, and data base accessible to practitioners, researchers, policy makers, and the public. We seek to link practitioners and researchers in the shared goal of improving mental health care by encouraging evidence-based practice and training.**

**Key words:** psychotherapy, outcome research, clinical trials, clinical practice, clinical training. [*Clin Psychol Sci Prac* 7:243–258, 2000]

Since the early 1990s, the Society of Clinical Psychology (Division 12) has focused increased attention on two potent trends: (a) the movement toward evidence-based practice, and (b) the increasing pressure for accountability in both clinical practice and clinical science. In 1993, David Barlow, then president of the society, created a Task Force (TF) on the Promotion and Dissemination of Psychological Procedures. The agenda of the original TF encompassed both education and public policy. A primary goal was to inform various constituencies about the state

of the science regarding clinical interventions: “This task force was constituted to consider methods for educating clinical psychologists, third party payors, and the public about effective psychotherapies” (Task Force on Promotion and Dissemination of Psychological Procedures, 1995, p. 3). The education agenda extended to training programs in clinical psychology, especially at the predoctoral and internship levels (Crits-Christoph, Frank, Chambless, Brody, & Karp, 1995; see also Calhoun, Moras, Pilkonis, & Rehm, 1998).

The Task Force on Promotion and Dissemination of Psychological Procedures, later renamed the Task Force on Psychological Interventions, and now the standing Committee on Science and Practice (CSP; following a 1998 change in Society bylaws) did groundbreaking work, developing criteria for levels of empirical support, identifying relevant treatment outcome studies, weighing the evidence in those studies according to the TF criteria, and presenting lists of treatments meeting criteria for different levels of empirical support (Chambless et al., 1998; Chambless et al., 1996; Task Force on Promotion and Dissemination of Psychological Procedures, 1995). The group also developed resource lists of treatment manuals and training opportunities for empirically supported treatments (ESTs; Sanderson & Woody, 1995; Woody & Sanderson, 1998).<sup>1</sup>

Related task forces within the Society have also done valuable work. The Task Force on Treatments That Work produced a volume entitled *A Guide to Treatments That Work*, edited by Peter Nathan and Jack Gorman (1998). Other task forces have focused on treatments for special populations. As one example, the Task Force on Lifespan Interventions focused on the treatment outcome literature with children and adolescents and with geriatric patients. The review process for children and adolescents led to a special issue of the *Journal of Clinical Child Psychology* (Lon-

---

Address correspondence to John Weisz, Department of Psychology, Franz Hall, 405 Hilgard Avenue, Los Angeles, CA 90095–1563. Electronic mail may be sent to weisz@psych.ucla.edu.

igan & Elbert, 1998) entitled “Empirically Supported Psychosocial Interventions for Children.” The issue covered treatments for child and adolescent depression, phobias and anxiety disorders, autism, conduct problems and disorders, and attention deficit hyperactivity disorder. In the months ahead, a series of reports on empirically supported treatments in pediatric psychology will be published in the *Journal of Pediatric Psychology*, under the general editorship of Anthony Spirito. As for the other end of the age spectrum, an article summarizing work on psychological treatments for older adults was published in the *Journal of Mental Health and Aging* (Gatz et al., 1998). The empirically supported treatments (EST) initiative has also received attention in Canada, where the Section on Clinical Psychology of the Canadian Psychological Association constituted its own Task Force on Empirically Supported Treatments. A summary report from that TF (Hunsley, Dobson, Johnston, & Mikail, 1999) and a series of commentaries (see Pilkonis, 1999) recently appeared in *Canadian Psychology*.

Related activities have proliferated, although not all of them are linked directly to the precedents created by CSP and its predecessors. We will not attempt a comprehensive listing, given the wide range of initiatives currently active in the area of evidence-based practice, but some good examples are

- the Task Force on Psychological Intervention Guidelines established by the American Psychological Association in 1993 and chaired by David Barlow, which published the *Template for Developing Guidelines: Interventions for Mental Disorders and Psychosocial Aspects of Physical Disorders* in 1995;
- a special section on “Validated Treatments: Multiple Perspectives and Issues,” published in *Clinical Psychology: Science and Practice* (Kazdin, 1996);
- a special section on “Psychotherapy Integration and Empirically Validated Treatments,” published in the *Journal of Psychotherapy Integration* (Glass & Arnkoff, 1996);
- a special issue on “Outcome Assessment of Psychotherapy” published by the *American Psychologist* (VandenBos, 1996);
- a journal, *Evidence-Based Mental Health*, inaugurated by the BMJ Publishing Group in 1998, that compiles abstracts of recent data-based articles, together with reviews intended to be useful to a multidisciplinary audience;
- a special section on “Treatment Manuals in Clinical

Practice” published in *Clinical Psychology: Science and Practice* (Kazdin, 1998);

- a special section on “The Empirically Supported Treatments Controversy” published in *Psychotherapy Research* (Elliott, 1998);
- a special section on “Empirically Supported Psychological Therapies” published in the *Journal of Consulting and Clinical Psychology* (Kendall & Chambless, 1998) that draws on the foundation provided by the Division 12 task forces, offers new thinking about how to evaluate and classify treatments (Chambless & Hollon, 1998), and presents reviews of empirically tested treatments in multiple domains of psychological dysfunction;
- a review of psychological and pharmacological interventions for the treatment of pain, which applied the Task Force criteria to studies in the field, published in *Clinical Psychology Review* (Wilson & Gill, 1996);
- a review of medication clinical trials with children and adolescents, which applied the Task Force criteria to pharmacotherapy studies, published in *Psychiatric Clinics of North America: Annual of Drug Therapy* (Riddle, Subramaniam, & Walkup, 1998); and
- a special section on “Clinical Significance” published in the *Journal of Consulting and Clinical Psychology* (Kendall, 1999).

There is much to appreciate in the work of the various “first generation” task forces. The concrete products noted above, together with the ongoing exchanges of ideas—both pro and con—published in this forum and others, have drawn attention to weighty issues in our field and prompted close examination in both the practice and research communities. The initial task force work provides a good foundation for further efforts by the new CSP. In this article, we describe some of the lessons we have learned from the task force work thus far, discuss several principles currently driving our work, and describe some of our plans for moving ahead. Our purpose is to inform our colleagues in research and practice about issues being discussed within the CSP and to stimulate discussion and debate within the field about our philosophy, objectives, and methods. We want to have this take place while our work is in process, so that the process itself, and the eventual products, can be enriched by the ideas of our colleagues in practice and research. Our focus here is on what might be called “the three Rs” of the CSP work—review procedures, research quality, and relevance to clinical practice.

## First Principles

Before we address some of the current priorities of the work of the Committee on Science and Practice, we want to note certain principles that guide our work:

*Principle 1 (Why we do our work).* Clinical practice and clinical training should be accountable to the public, and this accountability should be based on empiricism. Accountability in this sense includes justifying a given treatment plan (and changes to the plan) on the basis of scientific evidence. Of course, as in any scientific enterprise, the state of the evidence on procedures and their impact is continually evolving. Accordingly, what we owe the public is an ongoing, good-faith effort to keep training and practice aligned with the current state of the evidence.

*Principle 2 (How we do our work).* We need procedures for review and synthesis of evidence that answer questions most relevant to the scientific standing and clinical utility of treatments and that are replicable across different reviewers. Ongoing attention must be paid to literature review, rules of evidence, and classification procedures to insure that the information generated conveys a fair picture of the empirical status of treatments and their clinical impact. In addition, replicability of procedures and findings across reviewers is a goal, both to ensure that results of the review process are not unduly influenced by reviewer characteristics or attitudes and that changes in the review products over time reflect genuine changes in the existing evidence. This goal is important to any effort to construct a cumulative data archive on procedures and their effects, and such an archive is a long-term objective of the CSP.

*Principle 3 (How the work can benefit research).* Current treatment outcome research methods answer a few questions very well, but we need enriched methods that can generate information about the impact (and limits) of interventions across a broad range of conditions and populations. We also need information about the causal mechanisms underlying change. We have seen from earlier TF work that specifying and publicizing a set of standards for good quality research on psychotherapy can improve the research being done. The most common current paradigms for treatment outcome research warrant scrutiny and elaboration in order to encompass questions that are currently not well addressed, especially those questions that bear directly on clinical practice. Particularly impor-

tant will be designs that clarify those conditions under which interventions do and do not work well and designs that can enrich our understanding of causal processes underlying treatment effects. The CSP aims to promote this process by noting limitations in current evidence and by proposing strategies for strengthening the research armamentarium.

*Principle 4 (How the work can benefit practice).* Special attention must be paid to the challenge of moving treatments from clinical trials into clinical training and practice contexts. Researchers need to learn as much as possible about the real (complex and often multidisciplinary) world of practice, so that their work can lead to interventions that fit well into clinical service contexts (see, e.g., Wade, Treat, & Stuart, 1998). Even when this is accomplished, the transition from research settings and conditions to training and practice conditions will involve shifts along multiple dimensions, shifts that are neither simple nor inexpensive. Even the process of communication between researchers and clinicians needs attention; many practitioners do not find treatment research as reported in scientific journals to be particularly useful to them, and original research articles are not rated high among their sources of information on treatment (Cohen, Sargent, & Sechrest, 1986; Morrow-Bradley & Elliot, 1986). Given these realities, development and design of treatments, dissemination of treatment outcome information, and incorporation of CSP findings into clinical training and practice are all processes worthy of study (and support) in their own right. The work of the CSP in summarizing the state of the evidence should be complemented by collaborative assistance to those who want to strengthen the connection between such evidence and clinical training and practice.

In harmony with these principles, we see three broad goals for the work of the CSP:

- A. Improve mental health care for the public by encouraging and supporting evidence-based practice and by fostering clinical training that emphasizes reliance on empirically supported assessment and treatment procedures.
- B. Improve the quality of treatment research by encouraging increased attention to real-world clinical practice conditions, more rigorous tests of treatment effects, enriched outcome assessment that extends beyond symptoms and diagnoses, innova-

tions in the search for moderators and mediators of change, creation and testing of novel paradigms for treatment delivery, and increased emphasis on assessment of external validity, clinical utility, and effectiveness in practice.

- C. Support research, practice, and the public, by providing ongoing information on the state of evidence, as derived from standardized review and data aggregation procedures, and made available through cumulative data archiving across studies and over time.

### THE (OTHER) THREE RS

With these four principles and three broad goals as a backdrop, we now turn to the “three Rs” referred to at the outset, that is, the three primary emphases around which the agenda for the current CSP work has been organized. A summary of elements in that agenda is presented in Table 1.

#### Review Procedures: Rules of Evidence and Classification of Treatments

Previous TFs paid close attention to the development of fair procedures for reviewing and evaluating evidence and classifying treatments into levels of empirical support. The process involved a massive effort, and the resulting review procedures and classification rules served the needs of the initial phase of TF work quite well. In particular, the earliest TF placed a high priority on publishing and publicizing the “rules of evidence” in order to make the process of identifying empirically supported treatments as transparent and open to debate as possible. Why then should we focus our current efforts on review procedures and classification rules? We do this as part of a natural evolution, aimed at meeting scientific standards of reliability and addressing the needs of the practice and consumer communities. Although there was enormous controversy with the initial publication of TF decision rules, the debate was essential, instructive, and healthy for the field. Indeed, we anticipate that procedures and classification rules will continue to be revisited, to be debated anew, and to evolve with each new iteration of the CSP, just as the state and standards of our science and the needs of the consumer and practice communities evolve. We continue to place a high priority on openness in our attempts to elucidate empirically supported treatments, as we codify both the rules of evidence and procedures for applying those rules.

**Table 1.** Three-Part Agenda for the Committee on Science and Practice

#### I. Review Procedures and Rules of Evidence

Standardize literature search and review procedures  
Standardize rules of evidence for treatment classification decisions  
Simplify classification, employing one level of empirical support  
Assess reliability of standardized procedures and classification decisions  
Develop CSP deliberation procedures for last-stage decision making  
Debate long-range concerns (e.g., what yardstick to employ, character of “the list”)

#### II. Research Quality

Weigh alternative paradigms for comparing and synthesizing treatment outcome findings  
Promote increased attention to quality of evidence in review and dissemination  
Encourage clinical trials for a broadened range of treatments (including nonbehavioral)  
Encourage attention to broader array of outcome dimensions (including life functioning)  
Encourage identification of outcome moderators, to establish the “effective range” of benefit  
Encourage identification of outcome mediators, to establish mechanisms of change  
Evaluate alternative paradigms for outcome research (including “statistical modeling”)  
Identify and address professional barriers to identification of beneficial treatments  
Link outcome data within an accessible archive

#### III. Relevance to Clinical Practice and the Public Interest

Improve dissemination to practitioners, trainers, researchers, and the public (e.g., web site)  
Promote research on the transition of treatments from lab to clinic  
Identify and address obstacles to the clinical use of empirically tested treatments  
Support efforts by practitioners and trainers to incorporate empirically based approaches  
Promote rapprochement between research and practice communities

In earlier work, reviews were carried out separately by different TF members for different problems and disorders, with each reviewer using the search procedures deemed most productive. The original objective was to produce exemplars of empirically supported treatments, not a comprehensive list. In that spirit, TF members began their process by focusing on familiar treatments for which evidence was readily available. The first report issued by the TF reflected the literature bearing on those familiar treatments, rather than beginning with one or more diagnoses or problem areas and examining all relevant interventions directed toward it (cf. the Cochrane Collaboration and its search strategies; Chalmers, Sackett, & Silagy, 1997). Given this process, the earlier TF never contended that its lists of treatments were exhaustive. However, concern that “the list” might be too limited or doctrinaire did lead to expansion of the search and review procedures (see Chambless et al., 1998, pp. 3–4).

*Standardizing Review Procedures.* Characteristics of the initial TF process suggest at least two important objectives for the CSP: (a) ensuring that our review and classification

procedures are as transparent as possible and (b) establishing the reliability of the review and classification system. Accordingly, some of the CSP work will be focused on development of standardized procedures for searching and reviewing the treatment outcome literature, aiming for uniformity of procedures across the various problem and disorder categories and reliability across reviewers. In this effort, we face the constraints of time and effort inherent in any “volunteer” effort. Nonetheless, given the attention that has come to be focused on “the list,” there is a strong press toward both standardization and greater comprehensiveness.

*Rules of Evidence.* We also focus attention on the rules of evidence associated with review and classification. In evaluating the published evidence on various treatments, the first wave of task force reviews employed criteria and standards that were both thoughtful and reasonable. However, some of those criteria and standards involved judgments that may require more standard referents for uniform application across diverse studies.

To illustrate how important concepts introduced by the TF may require attention in our current process, consider two basic terms that appear in the TF criteria (see Task Force on Promotion and Dissemination of Psychological Procedures, 1995; Chambless et al., 1998): “replications of the *same* treatment,” and “experiments using *good* design.” The logic that all studies cited as evidence in support of a particular treatment must have used the same treatment is clear and compelling; on the other hand, it is difficult to find many cases in the literature involving separate tests of *exactly* the same treatment. Treatment development researchers tend to modify their interventions across successive trials (e.g., altering the treatment model, adding new sessions with new conceptual content, or changing the format and participants, as when an individual treatment becomes a group treatment or adds a family component). Such innovation raises the question of how it should be determined whether the treatments tested in any two clinical trials constitute “the same” treatment, and it is possible that different reviewers might come to different conclusions on specific cases (see Weisz & Hawley, 1998). If the CSP were to adopt a very strict requirement that “the same treatment” means identical manualized procedures across studies, few (if any) treatments could claim support in more than a single study. But a very loose requirement—say, that two studies are accepted as tests of the same treatment if the treatment developer says

they are the same—would risk claims of replication that might be difficult to support objectively in some cases. Clearly, the CSP will need some standard for “same treatment” that falls between the two extremes illustrated here. It could be argued that the best approach would be to define “same” in terms of the core elements of a treatment most responsible for its effects. But, because we do not fully understand the mechanisms by which even the best-researched treatments produce their effects, it is hard to know which components of a full intervention are “active” or “core,” and which are peripheral or merely matters of style. This highlights the need for the CSP to serve a research-gadfly function (in addition to its role in reviewing and classifying existing evidence), in this case prodding researchers in the direction of mediation tests (see below).

The flip side of determining whether two treatments purported to be “the same” are similar enough is determining whether two treatments purported to be “different” do in fact differ meaningfully. The task here is to determine what constitutes a sufficiently meaningful difference between nominally distinct treatments to warrant separate attention in the review process, and the possibility of separate listings in a tally of supported interventions. Is a 16-session cognitive-behavioral therapy program a “different treatment” than a 10-session program that addresses four fewer coping skills but has two booster sessions? If Smith publishes a study purporting to show no effects of treatment X, but Jones, who originally developed treatment X, argues that the treatment tested by Smith was not *really* X, should Smith’s or Jones’s judgment prevail? The CSP will need a rationale and procedures for making such judgments as these.

Another example of terminology that may need further operationalization is found in the requirement that studies employ “good between group design” or “good [single case] experimental designs” (see Chambless, 1998; Task Force on Promotion and Dissemination of Psychological Procedures, 1995). It was important for the TF to focus attention on study quality, and not to treat all studies as equal in their relevance to treatment efficacy. However, there is a possibility that different reviewers may differ in their views as to what constitutes *good* design, and even which elements of good design must be present for a study to be considered acceptable supporting evidence. Must the design be prospective, with random assignment? Must outcomes have been assessed with objective (e.g., performance test, behavior observation) measures or through

informants blind to treatment condition, or are self-report measures acceptable? Must adherence to the manualized treatment protocol have been assessed, to insure that the treatment ostensibly tested was in fact the treatment that participants received? Must attrition have been low and similar across groups? Must studies include follow-up assessment, to show that treatment effects persisted beyond immediate post-treatment? And should studies be disqualified if their outcome assessment includes only one symptom (e.g., state anxiety) of a multisymptom complex or disorder (e.g., posttraumatic stress disorder)? Problems in any of these areas, and of course many others, could threaten study validity, but which problem, or combination of problems, should mean that a study cannot be accepted as evidence in support of a particular treatment? To the extent that such questions can be answered *a priori*, the review and classification process becomes more transparent and more likely to be reliable across reviewers.

To help maximize uniformity in the review and classification process across reviewers, we will build on the work of earlier TFs, and such other sources as Chambless and Hollon's (1998) proposed criteria for good design, by (a) weighing what criteria are needed to fairly evaluate evidence on treatments, (b) operationally defining these criteria, (c) incorporating these criteria into a coding manual, and (d) testing interrater reliability in the application of the manualized criteria.

*Criteria Used for Treatment Classification Versus "Technical Codes."* Our starting point will be those criteria employed previously, but we are likely to modify the criteria, possibly in significant ways. For example, we expect to identify two types of evaluation criteria: (1) criteria used to signify scientific confidence that a particular treatment produces genuine change in the problems it is designed to treat (e.g., including those elements of "good design" that are considered critical to a study's inclusion as acceptable evidence), and (2) technical codes, used to characterize the nature and quality of evidence on various treatments, but not used to classify treatments (e.g., findings on moderators of treatment outcome). A goal will be to make data from these technical codes available for a variety of analyses and reports on the state of the field, ideally (eventually) in a searchable data archive accessed via the world wide web. As the process and the resulting information become more differentiated, there may be a need for dissemination in different formats for different audiences. For example, findings from the technical codes, reflecting the state of

the science, may be reported in journal articles for an audience that includes the research community, whereas summary information on the list of empirically supported treatments may be presented in less detail and in outlets more accessible to the general public (e.g., web sites, the popular press). In addition, we hope to see published discussion of pros and cons, contraindications, limitations and contradictions in the evidence, and critiques of the CSP process and products—all healthy for our evolving process.

*Classification Scheme for "Empirical Support," and Movement Toward a Single Category.* Recognizing that the quality and extent of empirical support can vary across a broad spectrum, the previous TFs distinguished between treatments that were "well-established" and treatments that were "probably efficacious" (see definitions in Task Force on Promotion and Dissemination of Psychological Procedures, 1995; Chambless et al., 1998). This has already proven to be an important distinction, one that helps focus the attention of researchers and treatment developers on key issues that need to be addressed in testing treatments, and in distinguishing moderate scientific support from strong support. On the other hand, dividing empirically supported treatments into these two categories may have certain limitations and may produce certain unintended consequences. One limitation, for example, may be that this two-group distinction, or any other, is incapable of capturing the multifactorial character of variations in empirical support. That is, there are literally dozens of variables reflecting research design quality and strength of support, any of which may be important in its own way, and only a few of which can be incorporated into even the best of two-category classification systems. Current thinking in the CSP is that the task of evaluating research quality and strength of support is so multidimensional that it might be best accomplished through an ongoing series of analyses presented as complements to the list of empirically supported treatments. We recognize that any system for distinguishing "empirically supported" treatments from those without such support necessarily involves an arbitrary judgment as to how good the evidence must be to warrant the designation "empirically supported." Nonetheless, two categories may give the impression of greater certainty and precision than we actually have.

Our thinking about the structure of the classification system has also been influenced by concerns about dissemination. What we have concluded, from consumer

and practitioner feedback, is that the two-category system may not offer the clarity needed to meet the needs of potential clinical users, both practitioners and clients. The use of two categories that are both “empirically supported” but that differ in level of support appears to have left some potential users confused. Should they rely only on “well established” treatments, or are “probably efficacious” treatments established *enough* for clinical use? Is it risky to try “probably efficacious” treatments? Should clinical training programs be training students in both classes of treatments, or only those that are “well established”? In short, the two-category system appears to pose different problems for different constituents. The two categories are insufficient to satisfy researchers who are intimately familiar with the nuances in levels of empirical support, and they are too detailed to readily convey information to practitioners and consumers. Balancing the disparate needs of these groups remains an ongoing challenge.

Thus, for both scientific and consumer-related reasons, the CSP is now inclined to work toward a single list of empirically supported treatments. Of course, a decision to create a single list raises a host of questions about what criteria should be employed—should the new criteria for single list membership be more similar to those of the “well-established” list or the “probably efficacious” list, or should the criteria involve new dimensions and new considerations beyond those employed previously? We invite input from interested readers on these and other questions raised in this article.

*Standard Setting Versus “Grandparenting.”* One of the most difficult dilemmas we confront is actually a result of progress in the field. Methodology in treatment outcome research grows more sophisticated over time, and the CSP membership wants our review criteria to serve a standard-setting function, with rules of evidence based on the current state of the science. But the state of the science was very different in, say, the 1960s, when many well-known and influential treatment studies were published. Strict criteria for “empirical support” that are based on 1990s standards might well exclude many outcome studies from earlier decades, and thus exclude a significant number of treatments from the list of empirically supported treatments. As one simple example, many reviewers in 2000 might feel that a study purporting to test treatment Y cannot be considered an acceptable test if the authors do not present evidence from coded session videotapes demon-

strating close adherence to the treatment Y protocol. But video coding to show adherence was not a part of treatment outcome research in the 1960s.

For this and a variety of other potential criteria, we face this question: Should we “grandparent in” treatments whose supporting studies are older and thus fail to satisfy current, strict criteria, if the bulk of the evidence supports those treatments? If so, we will be agreeing *not* to apply the same criteria consistently across studies. But if we reject grandparenting, and apply the same criteria consistently, we risk dropping from the list of ESTs some treatments that may actually be quite beneficial. We would also risk an ever-changing list composed primarily of recent treatments, since these will have been tested with the most recent version of acceptable outcome methodology.

To avoid the extreme version of this risk while still maintaining standards, our current thinking is to aim for criteria that reflect the spirit of the most current scientific standards, but permit some latitude in how those standards are satisfied. Thus, for example, we are inclined to require evidence of adherence, but also to accept such evidence in multiple forms—not only coding of therapy tapes, but also supervision of the treatment by experts in the protocol, or the use of treatment delivery methods (e.g., videotape) that provide a high degree of built-in adherence. This illustrates our current “golden mean” strategy for balancing standard-setting against grandparenting. The strategy will of course be increasingly difficult to implement with each passing decade, as prevailing standards grow increasingly different from those of earlier eras.

*Procedures for Committee Deliberation.* Given the virtually infinite array of forms treatment outcome research may take, and the great diversity of possible findings across various treatment conditions and outcome measures, it is difficult to produce a standard set of coding, decision, and classification rules that will address all possible studies and outcomes. Accordingly, the application of standardized procedures will almost certainly need to be complemented by committee deliberations on the state of the evidence bearing on each treatment under consideration. Thus, we are developing plans for committee action after the standardized procedures have been applied. A likely sequence is that subcommittees will apply review procedures and rules of evidence to treatments in their domains of interest, then present reports to the full CSP membership, with discussion of the reports leading to a formal vote by the CSP regarding each treatment proposed by

consideration. Our current thinking is that the proceedings of these meetings, summaries of the discussions, and results of the votes, should be made accessible to the public, through such outlets as periodic “update” articles, and entries on a web page (see below).

*Long-Range Concerns About “the Yardstick” and “the List.”*

Looking to the more distant future, we need to consider the yardstick by which to measure new candidates for inclusion on the list of empirically supported treatments, and impact of that yardstick on the nature of the list. One option is to continue adding treatments to the list based on evidence that they outperform only placebo or waitlist conditions (see Chambless et al., 1998). The findings of most meta-analyses with adults (e.g., Smith & Glass, 1977; Smith, Glass, & Miller, 1980) and children (e.g., Casey & Berman, 1985; Weisz et al., 1987, 1995b) suggest that most therapies tested in the published literature may be able to meet that standard. This in turn suggests the possibility of a very long list, over time. On the other hand, if the purpose of the list is to identify what works *best*, then comparative research becomes important, with designs in which new treatments are compared, not to a waitlist condition, but rather to the best available alternative treatment.

Adopting this emphasis would mean movement toward a list that rank-orders treatments for each target condition, with favorite oxen being gored from time to time, and with any current “best” treatment for condition X risking demotion as new upstarts prove more efficacious in new randomized controlled trials (RCTs). Of course, this dynamic process may be viewed as healthy for the field, and in fact analogous to the process by which lists of preferred antibiotics shift as new evidence accumulates. To be sure, the comparative process in psychotherapy would highlight a host of methodological issues. For example, when some relevant outcome measures favor treatment A and others treatment B, how shall we determine which measures constitute the most appropriate basis for determining rank? And suppose, as seems likely, that moderators of outcome will differ from one treatment to another, such that one treatment for condition X may work best for one subset of the population and another treatment may work best for another subset. Should cost effectiveness enter into a ranking process? Even if treatment A is somewhat less efficacious than treatment B across most outcome measures, it may still be seen as a

worthwhile candidate (at least from the payor’s perspective) if its cost is much lower. And should ranking be influenced by readiness of various treatments for use in clinical practice, with hard-to-incorporate treatments losing points in the process? The broad question here is whether a rank-ordering should be imposed over such variations, or whether, instead, we should scrap the idea of ranking and simply describe in detail, for each treatment, the conditions, groups, and other factors that define the costs, benefits, and effective range of that treatment? These questions are germane to the shape and character of “the list” and other products generated by the CSP, and are likely to fuel ongoing discussion and debate.

**Research Quality: Improving the Empirical Base**

Because of its role in reviewing evidence and evaluating treatments, the CSP is in a position to observe and comment on the state of outcome research, to note gaps and weaknesses in the empirical base, and to suggest strategies for addressing those gaps and weaknesses. We want to draw from our experience and perspective to raise issues that matter to the field, and help strengthen the scientific base of the discipline. By way of illustration, we highlight some of the issues here.

*Evolving Paradigms for Comparing and Synthesizing Treatment Outcome Studies.*

A fact of life in treatment outcome research is that paradigms evolve. For obvious reasons, we have a special interest in paradigms for identifying beneficial treatments (see Pilkonis, 1999). Since the late 1970s, a major influence on psychosocial treatment outcome research in the United States has been the regulatory model of the Food and Drug Administration (FDA). Within this model, the goal for any treatment is to establish evidence of its safety and statistically significant indications of its efficacy in at least two clinical trials. What this approach does not address is the quality of evidence overall. If two well-controlled studies support the efficacy of an intervention, how should one evaluate additional studies that do not show beneficial effects of the same treatment? Meta-analytic approaches address this issue, but the data base is constantly changing, and the effect size summaries open to alternate interpretations. For example, some evidence suggests that researcher allegiance is correlated with the magnitude of effect sizes (see Luborsky et al., 1999, for a discussion, with commentaries in the

same journal; but see Shirk & Russell, 1992, and Weiss & Weisz, 1995, for mixed evidence on this issue in child treatment research). If this is the case, then how should we weight supportive evidence from the treatment developers (or their progeny) relative to disconfirming evidence from independent or less invested researchers? It has also been observed that effect sizes decrease over time after treatment ends (e.g., Gaffan, Tsaousis, & Kemp-Wheeler, 1995; but see contrary evidence in Weisz, Weiss, Han, Granger, & Morton, 1995b). If this is the case, then what relative weight should be assigned to evidence of treatment effects immediately after treatment versus evidence of little effect at follow-up assessments six months or a year later? Another limitation of current approaches is that they only require that treatment conditions produce outcomes that are significantly better than control conditions. This ignores the question of whether such a statistical difference is clinically significant (Jacobson, Follette, & Revenstorf, 1984; Jacobson & Truax, 1991) and likely to have a significant impact on the quality of the client's life.

*Broadening the Range of Treatments Addressed and Outcomes Assessed.* Much of the research on treatment outcome is focused on behavioral (including cognitive behavioral) treatments, leaving our field relatively uninformed by empirical tests of many psychodynamic and other nonbehavioral psychotherapies that are widely practiced in real-world clinical settings. Well-designed empirical tests of the psychotherapies most often used in clinical practice would add importantly to the existing empirical base and inform practitioners of the evidence bearing on those treatments most relevant to them.

Among the empirical tests that have addressed treatment outcome, for either behavioral or other treatments, there is a heavy emphasis on outcome assessment of symptoms and diagnostic status, with less attention to such important areas as relationships with peers and loved ones, functioning in school and job settings, and future use of mental health and primary care services—all of which may be more important to treated clients and their significant others than diagnostic status. The narrow range of outcome assessment has come about in part because of the enormous influence of the diagnostic manuals published by the American Psychiatric Association, especially the *DSM-III* (American Psychiatric Association, 1980) and its successors. The influence has been magnified by the linkage of formal *DSM* diagnoses to entry into mental health

services and to reimbursement for such services. As psychotherapy research has been increasingly organized around *DSM* diagnostic categories, critics (e.g., Follette & Houts, 1996; Persons, 1986) have emphasized what we lose in the process—for example, information about psychological variables, about the real-world functioning of our clients, and about their perspectives on the treatments they have received. To enrich the picture of treatment outcome, we think it will be important for investigators to assess and to report analyses for an expanded range of outcome indicators (e.g., diagnostic status, target symptoms, general functioning, client satisfaction, subsequent service use; see Hoagwood, Jensen, Petti, & Burns, 1996; Jensen, Hoagwood, & Petti, 1996). Naturally, reporting more diverse indicators of outcome will highlight the issue of which indicators are most critical to the evaluation of treatment effects, and which should be used to determine membership on the list. As noted above, this is an issue we already face to some degree; having the issue grow more complex would be a small price to pay for the advantages of enriched outcome assessment.

*Testing the “Effective Range” of Treatment Benefit.* We also note that outcome assessment tends to be relatively short-term, with only a minority of studies reporting delayed follow-up assessments. Even in the child and adolescent treatment outcome studies where follow-up data are provided, the average length of the follow-up period is six months (Weisz & Weiss, 1993). Obviously, better information is needed on the staying power of intervention effects. The literature to date also provides relatively little evidence on moderators of treatment impact (see Beutler et al., 1991). We know relatively little, for example, about the extent to which ethnicity relates to the effects of various treatments that are currently on “the list,” either for adults (see Sue, 1995; Sue, Zane, & Young, 1994) or children and adolescents (see Weisz, Huey, & Weersing, 1998). Certainly we need to encourage research that will help identify the effective range within which empirically supported interventions can be expected to work, and outside of which they may be less useful or even contraindicated.

*Searching for Mechanisms of Change.* Of particular interest to the CSP is the goal of identifying the mechanisms through which efficacious treatments produce benefit. This goal is worthwhile both scientifically and practically.

Scientifically, intervention research needs to go beyond merely producing change, toward a causal account of the processes through which change occurs. Practically, efforts to develop interventions that are both efficacious and efficient can be guided by an understanding of which processes need to be set in motion to effect desired outcomes. Developments in statistical methodology for testing mediational models (Baron & Kenny, 1986; Shadish & Sweeney, 1991) are helpful in this regard, and mediation tests are now appearing in our best journals (see, e.g., Huey, Henggeler, Brondino, & Pickrel, in press). New thinking about designs for testing mediation hypotheses in treatment outcome research will be particularly useful (Follette, 1995; Follette & Compton, 1999). For example, if investigators are willing to specify the mechanism that is putatively responsible for producing change, and if they repeatedly measure hypothesized mediators throughout treatment, they can evaluate competing hypotheses as to why change occurs in the treatment group and locate the key components of the treatment while testing the accuracy of their treatment model. This approach can, in principle, help us distinguish between truly different therapies (see above) while advancing theory faster than is currently possible using traditional group design approaches. In this way, identifying mediators may help move us toward the worthwhile long-term goal of a parsimonious list, one that includes only those treatment components that independently or synergistically contribute to change.

*Evaluating Alternative Paradigms for Psychotherapy Outcome Research.* Different research paradigms have different utilities, but many different paradigms have the potential to add meaningfully to our understanding of the process and outcome of psychotherapy. For example, in a “component analysis” paradigm, the focus is on mechanisms of change and the goal is to examine one treatment package and then, with increasing precision, to isolate the active ingredients of this treatment (Follette & Compton, 1999). Yet another paradigm is the “statistical modeling” paradigm. Howard and colleagues (Howard, Moras, Brill, Martinovich, & Lutz, 1996) have made a particularly strong case for advancing knowledge on the basis of modeling large numbers of realized cases and not “overvaluing” randomized controlled trials (RCTs). According to Howard et al., placing too high a premium on RCTs has the consequence of stating the results of experiments

involving random assignment too confidently (they are only quasi-experiments after all because of what is not randomized) and treating “correlational studies” as if they were qualitatively worse than such experiments (whereas the issue is a quantitative one of more versus fewer threats to internal validity). A related goal here is to produce research results that clinicians will eventually be able to apply to individual cases—a “quality control” paradigm in which clinicians can monitor the outcomes of their patients, within reliable confidence intervals, and make adjustments in real time when results are falling below expected levels.

*Noting Professional Barriers and Facilitators.* There may also be merit in attending to current practices in our profession that impede or might facilitate the process of identifying empirically supported treatments. The editorial policy of major journals, for example, could do a great deal to promote the process. A major change that would bring major benefit would be journal policy requiring uniform reporting of those aspects of study design (e.g., method of assignment to groups), methodology (e.g., whether adherence checks were carried out, and showed that the treatment was delivered as intended), and findings (e.g., means and SDs for each group on each outcome measure at each point of assessment) needed for thorough review and synthesis of multiple studies. As another illustration, if editors would consistently enforce the APA journal policy that effect sizes be reported in treatment outcome studies, this would greatly facilitate the process of weighing evidence across studies. As things stand now, studies with abundant statistical power can produce statistically significant treatment effects despite very modest treatment-versus-control group differences, and two treatments producing statistically significant effects are regarded as equivalent even though their effect magnitude may differ greatly. Uniform application of an effect size metric could go a long way toward generating meaningful comparison of benefit across studies and treatments.

There are additional professional impediments to the identification of evidence-based treatment. For example, the premium that the work of the CSP places on replication of outcome findings conflicts with the emphasis in our discipline on originality of contributions. Many decisions in academia, ranging from hiring to publication to grant funding, are made in favor of original contributions

over repeat performances and retests of other investigators' (or one's own) findings. The CSP work has rightly stressed our need to know whether outcome findings are robust or mere flashes in the pan, but to satisfy this need, our field must find ways to encourage, finance, and reward carefully done replications. Otherwise, the incentive system in the discipline may continue to work against an important form of research that remains essential to full evaluation of treatments.

*Linking Outcome Data Within an Accessible Archive.* The push toward evidence-based practice would be fundamentally enriched by a much-needed, but difficult to achieve, development to which we alluded earlier: creation of a data archive, linking treatment outcome studies across time and setting. Our work increasingly persuades us of (a) the value of integrating findings from multiple studies, and (b) the difficulty of such integration, given the heterogeneity of reporting conventions followed by various authors. To support efforts at integration of findings across studies, we encourage the development of a treatment outcome data archive, guided by a template that will encourage uniformity in important dimensions of data reporting, including study design, features of the sample, a summary of all data collected, and the nature and results of data analyses. The obstacles to such an enterprise are massive, and the startup cost could be substantial. Yet we need some mechanisms for aggregation of results across studies, if our science is to achieve the incremental, cumulative process most of us believe is critical. Such aggregation would also facilitate communication of research findings directly to practitioners and consumers.

**Relevance to Clinical Practice: Dissemination, Outreach, and Transportability**

Why focus on relevance to clinical practice? One major objective of the original TF was to identify treatments whose level of empirical support warranted their use as first-line approaches in clinical training and practice, and to educate clinicians, clients, and policy makers about their availability. However, we have learned in the interim that simply identifying ESTs does not guarantee that they will be adopted and used in clinical practice settings, requested by consumers, or embraced by managed systems of health care. Moreover, the various ESTs may well

differ in the ease with which they can be applied in practice contexts. These ideas lead us in several directions.

*Improving Dissemination and Outreach: How to Get the Word Out?* One such direction involves dissemination and outreach. We are currently considering what may be the best ways to inform the public, the practitioner community, and other interested parties, about the ESTs identified, and about procedures for accessing more information. Individuals in the general public will need to know where and how they may gain access to one of these treatments. Interested practitioners will need information on how they may obtain written protocols and training in how to use them. Managed care will need information on the cost and length of various treatments, the likelihood of benefit for a given patient, and the disorders for which an empirically supported treatment exists. In a discipline not known for elaborate self-promotion or commercial advertising (in contrast to, say, the pharmaceutical industry), we must learn the best ways to reach various constituencies.

And we must learn to do so efficiently. Managed care, governmental bodies, and consumer groups are prone to make policy decisions regarding mental health and other matters in relatively quick flurries of activity, often spurred on by a major social event or a recent headline; by contrast, the TF/CSP process of review and synthesis of the evidence is deliberative and, well, slow. One of our challenges will be to make the information needed by various constituents available in a timely way, so that those developing policies that bear on mental health care, and facing time pressure in the process, will have the evidence they need when they need it. If the information needed happens to match what we have already assembled, or can be readily extracted from the database discussed above, then there may be no problem. But it remains to be seen whether the CSP can be nimble enough to meet most agency and consumer group demands in a timely fashion.

*Using the Web.* Thinking about nimbleness brings us to the theme of electronic communication. As more clinicians and consumers gain access to the Internet, the worldwide web is becoming an increasingly feasible way to publish information that requires periodic modification and updating. The CSP is currently developing a web site (via <http://www.apa.org/divisions/div12>) that will pro-

## Depression

Table of Contents
Introduction
Anxiety Disorders and Stress
Borderline Personality Disorder
Childhood Disorders
Depression
Drug and Alcohol Abuse
Eating Disorders
Hair Pulling
Health-Related Problems
Marital Distress
Schizophrenia and Other Severe Mental Illness
Sexual Dysfunctions

**Figure 1.** Sample portion of the draft web page being developed by the Committee on Science and Practice. To review the entire web site, go to <http://www.apa.org/divisions/div12/homepage.shtml>

### Disclaimer

Clinical depression is more than just feeling blue or down after a bad day at work. People who are clinically depressed feel down, sad, or hopeless all the time, for weeks on end. They often become disinterested in things they used to enjoy. In addition to feeling low all the time, depressed persons often have trouble sleeping or eating. They find that they have trouble with their thinking; they may not be able to concentrate well enough to read or even watch television. Depressed persons often spend a lot of time thinking about death, or thinking that they would be better off dead. These symptoms are similar to those experienced by someone who is grieving the death of a loved one. However, in depression, these feelings arise without such a loss, or they last much longer than normal grief.

Behavior therapy, cognitive therapy, and interpersonal therapy have all been well-established as beneficial treatments for major depression. In addition, some evidence suggests that **brief dynamic therapy**, **self-control therapy**, and **social problem-solving therapy** are useful in the treatment of major depression. Finally, some evidence from studies with older adults suggests that cognitive therapy and **reminiscence therapy** are useful in the treatment of geriatric major depression. While other psychotherapies may be helpful in the treatment of depression, they have not been evaluated scientifically in the same way as the treatments listed here. Many medications are also helpful for depression, but we do not cover medications in this website. Of course, we recommend a consultation with a mental health professional for an accurate diagnosis and discussion of various treatment options. When you meet with a professional, be sure to work together to establish clear treatment goals and to monitor progress toward those goals. Feel free to print this information and take it with you to discuss your treatment plan with your therapist.

For more information about depression, advice on obtaining treatment, and support from other sufferers, click on Wing of Madness: A Depression Guide. Another informative site with information on depression symptoms, treatment, and online resources is maintained by Mental Health Net.

vide the basic information presented through previous TF reports and future CSP reports. As the sample draft page shown in Figure 1, illustrates, we are aiming for jargon-free, accessible language in reporting on problems and associated treatments. The site is organized around problem areas, with links to other Internet sites that provide (in our judgment) high-quality information about each disorder or problem. The CSP site provides a few sentences about each problem area, and users can access more detailed information by clicking on the link to other sites. The CSP site also provides information about which treatments have empirical support for each problem, along with an explanation of what “empirical support” means. In many cases, there is a description of the supported treatment, provided in sufficient detail to permit consumers to judge whether a treatment they are currently receiving resembles a tested intervention. The site

information also encourages consumers to print out the information and discuss it with their practitioners. A working version of the site is available through the Society of Clinical Psychology (APA Division 12) web site [go to [http://www.apa.org/divisions/div12/rev\\_est/index.shtml](http://www.apa.org/divisions/div12/rev_est/index.shtml)].

*Addressing Barriers Faced by Practitioners in the Use of Evidence-Based Procedures.* Even if abundant information were available, through the web and other media, practitioners would still face numerous professional barriers to learning and implementing evidence-based practices. One of the biggest impediments may be the lack of financial support for assessment. Without reliable and valid initial client assessment, the prospects for identifying appropriate empirically supported treatment are undermined. And without ongoing (reliable and valid) measurement of

progress, there can be little hope of implementing the kinds of actuarial approaches to treatment planning advocated by Howard et al. (1996). An additional impediment is that there are few professional incentives for learning new treatment approaches, and with ever-increasing case loads and “productivity monitoring” in many service settings, the time needed for new learning is in ever-shorter supply. Although many states require continuing education for practice licensure, there is no requirement that any of this education be related to empirically supported treatments. Furthermore, for many of these treatments, simply attending a continuing education workshop does not provide sufficient expertise to employ the treatment in practice. Instead, postworkshop supervision is needed, and such supervision is often difficult to arrange, sometimes expensive, and typically the financial responsibility of the individual clinician. In sum, those practitioners who are not currently using empirically supported procedures, but would like to do so, may face significant obstacles. We need to find ways to support their efforts.

By educating academics about the challenges faced by clinicians, we also hope to influence researchers to include design elements that facilitate the transfer of interventions from the laboratory to the consulting room. Such design elements would include cost-effectiveness analysis, not just of the intervention, but also of various assessment components, a broader perspective on measurement of outcomes, and summary data presented in ways that are meaningful to clinicians (see Woody & Kihlstrom, 1997).

*Addressing Issues of Transportability.* A second path leads us to think about characteristics of empirically supported treatment procedures themselves, and the research contexts in which they have been tested, that may limit the ease with which they can be incorporated into practice and training. The underlying objective is to help, wherever possible, to facilitate the use of empirically supported treatments as an essential component of quality mental health care. Increasingly, we think it will be useful for CSP to focus attention on the readiness of such treatments for use in practice. There are multiple dimensions (e.g., client heterogeneity, comorbidity, and severity; therapist training and caseloads; infrastructure support for standardized procedures) along which treatment in most RCTs differs from treatment in practice settings (see Weisz, Donenberg, Han, & Weiss, 1995). We certainly believe that

empirically supported treatments are more likely to be beneficial than untested treatments, not only in clinical trials but also in clinical practice. But recognition of the numerous differences between treatment in trials and treatment in practice does highlight the potential difficulty of transporting lab-derived treatments into practice settings. All this suggests a need for a genre of research focused on the transition from lab to clinic. Research within this genre might address such questions as what adaptations are needed to make standardized treatments workable and effective with complex clients who have multiple diagnoses and who (because they are not study volunteers) may not want to do out-of-session homework, what training and supervision procedures can inculcate new expertise within the time constraints of a busy clinical practice, and what clinic infrastructure changes are required to make effective use of empirically supported treatments a routine part of everyday practice. Particularly important will be tests of the empirically supported treatments in clinical practice contexts and conditions. We will aim to provide information to potential users on the extent to which such research has been done, and what it has shown, by including information along these lines in the CSP coding system. As examples, we may code outcome studies for the degree to which the clients treated in those studies are or resemble clients seen in practice settings (e.g., clinically referred, diagnosed, and with comorbid conditions), the degree to which the treatments have been administered by practitioners (versus graduate students in the treatment developer’s lab), the degree to which the treatment has been given in a clinical service setting rather than a primarily research setting, and so forth. In addition, by focusing attention on the issue of practice-relevant testing, through our technical codes, we hope to promote outcome research that is done under conditions that resemble clinical practice more than the traditional RCT.

*Promoting Rapprochement Between the Research and Practice Communities.* There is also a need to address the tension between the research and clinical practice communities, which certainly constitutes an obstacle. To reduce the ceaseless casting of blame between researchers and clinicians, we need a mutual rapprochement, and this may be fostered by a recognition of what each group might learn from and contribute to the other. Researchers must learn more about the realities of practice if their work is to lead

to products that are maximally relevant to real-world treatment. The increased NIMH emphasis on effectiveness, practice, and service systems research in the United States is pushing researchers in this direction (Clinical Treatment and Services Research Workgroup, 1998). The push demands more research attention to clinical epidemiology (i.e., the study of who actually appears for which treatments in which settings), to organizational issues at the level of practice groups and service delivery systems, and to factors in practice that merit replication in research tests of treatment outcome. On the other hand, clinicians might consider the potentially important organizing influences that best research practices can provide. In our view, "usual care" can be improved not only by the use of evidence-based procedures, but also by greater attention to systematic assessment, organized attempts to link assessment to treatment, ongoing enhancement of clinical skills through education about empirical findings, and longitudinal evaluation, all done in the spirit of accountability. In our specialized world, practitioners and researchers have distinctive experience, knowledge, and skills, and each group can learn a good deal from the other. Indeed, we suspect that collaboration will bring out the best in both groups, and to the benefit of those who seek mental health care.

#### NOTE

1. Several of these reports are available at two sites on the World Wide Web: the web page of Division 12 at <http://www.apa.org/divisions/div12/est/est.htm> and the web page of the Society for a Science of Clinical Psychology (Section III of Division 12) at [http://www.sscp.psych.ndsu.nodak.edu/sscpweb/est\\_docs/tf\\_docs.htm](http://www.sscp.psych.ndsu.nodak.edu/sscpweb/est_docs/tf_docs.htm)

#### ACKNOWLEDGMENTS

William Follette, Paul Pilkonis, John Weisz, and Sheila Woody serve on the Committee on Science and Practice (CSP) of the Society of Clinical Psychology (Division 12 of the American Psychological Association), and Kristin Hawley serves as an advisor to the CSP. We appreciate feedback from various members of the CSP, and we particularly want to thank Naihua Duan for his thoughtful suggestions regarding the manuscript.

#### REFERENCES

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders*. (3rd ed.). Washington, DC: Author.  
 Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Concep-

tual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.  
 Beutler, L. E., Engle, D., Mohr, D., Daldrup, R. J., Bergan, J., Meredith, K., Merry, W. (1991). Predictors of differential response to cognitive, experiential, and self-directed psychotherapeutic procedures. *Journal of Consulting and Clinical Psychology*, *59*, 333-340.  
 Calhoun, K. S., Moras, K., Pilkonis, P. A., & Rehm, L. P. (1998). Empirically supported treatments: Implications for training. *Journal of Consulting and Clinical Psychology*, *66*, 151-162.  
 Casey, R. J., & Berman, J. S. (1985). The outcome of psychotherapy with children. *Psychological Bulletin*, *98*, 388-400.  
 Chalmers, I., Sackett, D., & Silagy, C. (1997). The Cochrane Collaboration. In A. Maynard & I. Chalmers (Eds.), *Non-random reflections on health services research: On the 25th anniversary of Archie Cochrane's "Effectiveness and efficiency"* (pp. 231-249). London: BMJ Books.  
 Chambless, D. L., Baker, M. J., Baucom, D. H., Beutler, L., Calhoun, K. S., Crits-Christoph, P., Daiuto, A., DeRubeis, R., Detweiler, J., Haaga, D. A. F., Johnson, S. B., McCurry, S., Mueser, K. T., Pope, K. S., Sanderson, W. C., Shoham, V., Stickle, T., Williams, D. A., & Woody, S. A. (1998). Update on empirically validated therapies, II. *The Clinical Psychologist*, *51*, 3-16.  
 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., Bennett Johnson, S., 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. *The Clinical Psychologist*, *49*, 5-18.  
 Clinical Treatment and Services Research Workgroup. (1998). *Bridging science and service*. Washington, DC: National Institute of Mental Health.  
 Cohen, L., Sargent, M., & Sechrest, L. (1986). Use of psychotherapy research by professional psychologists. *American Psychologist*, *41*, 198-206.  
 Crits-Christoph, P., Frank, E., Chambless, D. L., Brody, C., & Karp, J. F. (1995). Training in empirically validated treatments: What are clinical psychology students learning? *Professional Psychology: Research and Practice*, *26*, 514-522.  
 Elliott, R. (Ed.). (1998). The empirically supported treatments controversy [special section]. *Psychotherapy Research*, *8*, 115-170.  
 Follette, W. C. (1995). Correcting methodological weaknesses in the knowledge base used to derive practice standards. In S. C. Hayes, V. M. Follette, R. M. Dawes, & K. E. Grady (Eds.), *Scientific standards of psychological practice: Issues and recommendations* (pp. 229-247). Reno, NV: Context Press.

- Follette, W. C., & Compton, S. N. (1999). *Correcting methodological weaknesses in traditional psychotherapy outcome research*. Unpublished manuscript.
- Follette, W. C., & Houts, A. C. (1996). Models of scientific progress and the role of theory in taxonomy development: A case study of the DSM. *Journal of Consulting and Clinical Psychology, 64*, 1120–1132.
- Gaffan, E. A., Tsaousis, J., & Kemp-Wheeler, S. M. (1995). Researcher allegiance and meta-analysis: The case of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 63*, 966–980.
- Gatz, M., Fiske, A., Fox, L. S., Kaskie, B., Kasl-Godley, J. E., McCallum, T. J., & Wetherell, J. L. (1998). Empirically validated psychological treatments for older adults. *Journal of Mental Health and Aging, 4*, 9–46.
- Glass, C. R., & Arnkoff, D. B. (Eds.). (1996). Psychotherapy integration and empirically validated treatments [special section]. *Journal of Psychotherapy Integration, 6*, 183–226.
- Hoagwood, K., Jensen, P. S., Petti, T., & Burns, B. J. (1996). Outcomes of mental health care for children and adolescents: I. A comprehensive conceptual model. *Journal of the American Academy of Child and Adolescent Psychiatry, 35*, 1055–1063.
- Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z., & Lutz, W. (1996). Evaluation of psychotherapy: Efficacy, effectiveness, and patient progress. *American Psychologist, 51*, 1059–1064.
- Huey, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (in press). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through improved family and peer functioning. *Journal of Consulting and Clinical Psychology*.
- Hunsley, J., Dobson, K., Johnston, C., & Mikail, S. (1999). Empirically supported treatments in psychology: Implications for Canadian professional psychology. *Canadian Psychology, 40*, 289–302.
- Jacobson, N. S., Follette, W. C., & Revenstorff, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15*, 336–352.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.
- Jensen, P. S., Hoagwood, K., Petti, T. (1996) Outcomes of care for children and adolescents: II. Application of the SFCEs model to current research. *Journal of the American Academy of Child and Adolescent Psychiatry, 35*, 1064–1077.
- Kazdin, A. E. (Ed.). (1996). Validated treatments: Multiple perspectives and issues [special section]. *Clinical Psychology: Science and Practice, 3*, 216–267.
- Kazdin, A. E. (Ed.). (1998). Treatment manuals in clinical practice [special section]. *Clinical Psychology: Science and Practice, 5*, 361–407.
- Kendall, P. C. (Ed.). (1999). Clinical significance [special section]. *Journal of Consulting and Clinical Psychology, 67*, 283–339.
- Kendall, P. C., & Chambless, D. L. (Eds.). (1998). Empirically supported psychological therapies [special section]. *Journal of Consulting and Clinical Psychology, 66*, 3–167.
- Lonigan, C. J., & Elbert, J. C. (Eds.). (1998). Empirically supported psychosocial interventions for children [Special issue]. *Journal of Clinical Child Psychology, 27*(2).
- Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., Halperin, G., Bishop, M., Berman, J. S., & 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.
- Morrow-Bradley, C., & Elliott, R. (1986). Utilization of psychotherapy research by practicing psychotherapists. *American Psychologist, 41*, 188–197.
- Nathan, P. E., & Gorman, J. M. (Eds.). (1998). *A guide to treatments that work*. New York: Oxford University Press.
- Persons, J. B. (1986). The advantages of studying psychological phenomena rather than psychiatric diagnoses. *American Psychologist, 41*, 1252–1260.
- Pilkonis, P. A. (Ed.). (1999). Special section: Paradigms for psychotherapy outcome research. *Journal of Clinical Psychology, 55*, 145–205.
- Pilkonis, P. A. (1999). Let us move ahead, together: Commentary on "Empirically supported treatments in psychology: Implications for Canadian professional psychology." *Canadian Psychology, 40*, 303–306.
- Riddle, M. A., Subramaniam, G., & Walkup, J. T. (1998). Efficacy of psychiatric medications in children and adolescents: A review of controlled studies. *Psychiatric Clinics of North America: Annual of Drug Therapy, 5*, 269–285.
- Sanderson, W. C., & Woody, S. R. (1995). Manuals for empirically validated treatments: A project of the Task Force on Psychological Interventions. *The Clinical Psychologist, 48*(4), 7–11.
- Shadish, W. R., & Sweeney, R. B. (1991). Mediators and moderators in meta-analysis: There's a reason why we don't let dodo birds tell us which psychotherapies should have prizes. *Journal of Consulting and Clinical Psychology, 59*(6), 883–893.
- Shirk, S. R., & Russell, R. L. (1992). A reevaluation of estimates of child therapy effectiveness. *Journal American Academy of Child and Adolescent Psychiatry, 31*, 703–709.
- Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*, 752–760.
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.

- Sue, S. (1995). The implications of diversity for scientific standards of practice. In S. C. Hayes, V. M. Follette, R. M. Dawes, & K. E. Grady (Eds.), *Scientific standards of psychological practice: Issues and recommendations* (pp. 265–279). Reno, NV: Context Press.
- Sue, S., Zane, N., & Young, K. (1994). Research on psychotherapy with culturally diverse populations. In A. E. Bergin & S. L. Grady (Eds.), *Handbook of psychotherapy and behavior change* (pp. 783–817). New York: Wiley.
- Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically-validated psychological treatments. *The Clinical Psychologist*, 48, 3–23.
- Task Force on Psychological Intervention Guidelines. (1995). *Template for developing guidelines: Interventions for mental disorders and psychosocial aspects of physical disorders*. Washington, DC: American Psychological Association.
- VandenBos, G. R. (Ed.). (1996). Outcome assessment of psychotherapy [special issue]. *American Psychologist*, 51, 1005–1006.
- Wade, W. A., Treat, T. A., & Stuart, G. L. (1998). Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. *Journal of Consulting and Clinical Psychology*, 66, 231–239.
- Weiss, B., & Weisz, J. R. (1995). Effectiveness of psychotherapy. *Journal of the American Academy of Child and Adolescent Psychiatry*, 34, 971–972.
- Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995a). Bridging the gap between lab and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology*, 63, 688–701.
- Weisz, J. R., & Hawley, K. M. (1998). Finding, evaluating, refining, and applying empirically supported treatments for children and adolescents. *Journal of Clinical Child Psychology*, 27, 205–215.
- Weisz, J. R., Huey, S. J., & Weersing, V. R. (1998). Psychotherapy outcome research with children and adolescents: The state of the art. *Advances in Clinical Child Psychology*, 20, 49–91.
- Weisz, J. R., & Weiss, B. (1993). *Effects of psychotherapy with children and adolescents*. Newbury Park, CA: Sage.
- Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology*, 55, 542–549.
- Weisz, J. R., Weiss, B., Han, S., Granger, D. A., & Morton, T. (1995b). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin*, 117, 450–468.
- Wilson, J. L., & Gil, K. M. (1996). The efficacy of psychological and pharmacological interventions for the treatment of chronic disease-related and non-disease-related pain. *Clinical Psychology Review*, 16, 573–597.
- Woody, S. R., & Kihlstrom, L. C. (1997). Outcomes, quality, and cost: Integrating psychotherapy and mental health services research. *Psychotherapy Research*, 7, 365–382.
- Woody, S. R., & Sanderson, W. C. (1998). Manuals for empirically supported treatments: 1998 update. *The Clinical Psychologist*, 51, 17–211.

Received June 1, 1999; revised October 12, 1999; accepted October 21, 1999.