Psychological Bulletin
1997, Vol. 122, No. 3, 203-215

0033-2909/97/$3.00

A Meta-Analysis of Outcome Studies Comparing Bona Fide Psychotherapies: Empirically, "All Must Have Prizes"

Bruce E. Wampold, Gregory W. Mondin, Marcia Moody, Frederick Stich, Kurt Benson, and Hyun-nie Ahn
University of Wisconsin—Madison

This meta-analysis tested the Dodo bird conjecture, which states that when psychotherapies intended to be therapeutic are compared, the true differences among all such treatments are 0. Based on comparisons between treatments culled from 6 journals, it was found that the effect sizes were homogeneously distributed about 0, as was expected under the Dodo bird conjecture, and that under the most liberal assumptions, the upper bound of the true effect was about .20. Moreover, the effect sizes (a) were not related positively to publication date, indicating that improving research methods were not detecting effects, and (b) were not related to the similarity of the treatments, indicating that more dissimilar treatments did not produce larger effects, as would be expected if the Dodo bird conjecture was false. The evidence from these analyses supports the conjecture that the efficacy of bona fide treatments are roughly equivalent.

In 1936, Rosenzweig proposed that common factors were responsible for the efficacy of psychotherapy and used the conclusion of the Dodo bird from *Alice in Wonderland* (Carroll, 1865/1962) to emphasize this point: "At last the Dodo said, 'Everybody has won, and all must have prizes' " (p. 412). Later, Luborsky, Singer, and Luborsky (1975) reviewed the psychotherapy outcome literature, found that the psychotherapies reviewed were generally equivalent in terms of their outcomes, and decreed that the Dodo bird was correct. Since Luborsky et al.'s seminal review, the equivalence of outcome in psychotherapy has been called the *Dodo bird effect.*

To many interested in the technical aspects of particular psychotherapies, the Dodo bird effect was distasteful and, on the face of it, unbelievable:

> If the indiscriminate distribution of prizes argument carried true conviction . . . we end up with the same advice for everyone—"Regardless of the nature of your problem seek any form of psychotherapy." This is absurd. We doubt whether even the strongest advocates of the Dodo bird argument dispense this advice. (Rachman & Wilson, 1980, p. 167)

So the race has been run over and over again. In fact and maybe because of the Dodo bird conclusion, psychotherapy research has become increasingly pragmatic, designed to detect winners and losers (Omer & Dar, 1992). Lately, the races have been sanctioned by the psychotherapy community. In response to external pressures, Division 12 of the American Psychological Association (APA) proclaimed that "if clinical psychology is to survive in this heyday of biological psychiatry, APA must act to emphasize the strength of what we have to offer—a variety of psychotherapies of proven efficacy" (Task Force on Promotion and Dissemination of Psychological Procedures, 1995). Accordingly, criteria were developed to identify empirically validated treatments (Task Force on Promotion, 1995, p. 21); these criteria refer exclusively to studies that assess outcome rather than to process, theory, or psychological mechanisms of change, elevating the importance of winners (Wampold, 1997). In fact, 18 therapies have been identified as winners by the task force and consequently designated as empirically validated (Task Force on Promotion, 1995, Table 3).

Clearly, to identify the winners and losers among the set of treatments, the race should be fair. In this regard, the allusion to the race from *Alice in Wonderland* (Carroll, 1865/1962) was problematic because the competitors participated haphazardly:

> [The competitors] were placed along the course, here and there. There was no "One, two, three and away," but they began running when they liked, and left off when they liked so that it was not easy to know when the race was over. (p. 45)

To receive a prize that has meaning, the competitors must have a level playing field—to mix the metaphor slightly. Occasionally, individual races have produced winners (e.g., Butler, Fennell, Robson, & Gelder, 1991; Snyder, Wills, & Grady-Fletcher, 1991), but the loser in any such race can always find conditions that put them at a disadvantage. For example, when it was found that insight-oriented marital therapy (IOMT) resulted in dramatically fewer divorces than behavioral marital therapy (BMT) 4 years after therapy (3% and 38%, respectively; Snyder et al., 1991), Jacobson (1991) argued that IOMT had an unfair advantage:

> It seems obvious that the IOMT therapists were relying heavily on the nonspecific clinically sensitive interventions allowed in the

Bruce E. Wampold, Gregory W. Mondin, Marcia Moody, Frederick Stich, Kurt Benson, and Hyun-nie Ahn, Department of Counseling Psychology, School of Education, University of Wisconsin—Madison.

This research was supported, in part, by appointment of Bruce E. Wampold as a Vilas Associate in the Social Sciences at the University of Wisconsin—Madison.

Correspondence concerning this article should be addressed to Bruce E. Wampold, Department of Counseling Psychology, School of Education, University of Wisconsin, 321 Education Building, 1000 Bascom Mall, Madison, Wisconsin 53706. Electronic mail may be sent via Internet to wampold@macc.wisc.edu.

> IOMT manual but not mentioned in the BMT manual. . . . To me, the . . . data suggest that *in this study* BMT was practiced with insufficient attention to nonspecifics. (p. 143)

Even the most expensive and thoroughly conducted clinical trial ever conducted (viz., the NIMH [National Institute of Mental Health] Treatment of Depression Collaborative Research Program) could not inoculate the results against claims that a treatment was at a disadvantage for some reason (Elkin, Gibbons, Shea, & Shaw, 1996; Jacobson & Hollon, 1996a, 1996b; Klein, 1996).

Because every study has some flaws and because any single study that shows the superiority of a treatment could be due to a Type I error, statements about relative efficacy of treatments based on the results of a single study are unjustified (Wampold, 1997). Recognizing the limitations of making inferences from a single study, researchers have used meta-analytic methods to empirically examine the Dodo bird effect. Smith and Glass's (1977) seminal review of psychotherapy outcome studies was the first attempt to meta-analytically test whether any particular type of therapy was superior to another. For each study, they calculated the effect size of the psychotherapy vis-à-vis a control group. Then by aggregating these effect sizes within categories of therapies (e.g., Adlerian, systematic desensitization), Smith and Glass compared the relative size of the resultant effects and found small differences among the categories—about 10% of the variance in effect sizes was due to category. However, after aggregating into super classes and equating the classes for differences in studies, they found that the data yielded a result consistent with the Dodo bird effect: "Despite volumes devoted to the theoretical differences among different schools of psychotherapy, the results of research demonstrate negligible differences in the effects produced by different therapy types" (p. 760). On the basis of the Smith and Glass and subsequent meta-analyses that examined the relative efficacy of psychotherapies, Lambert and Bergin (1994) concluded that

> there is a strong trend toward no difference between techniques in amount of change produced, which is counterbalanced by indications that, under some circumstances, cognitive and behavioral methods are superior even though they do not generally differ in efficacy between themselves. An examination of selected exemplary studies allows us to further explore this matter. Research carried out with the intent of contrasting two or more bonafide treatments show surprisingly small differences between the outcomes for patients who undergo a treatment that is fully intended to be therapeutic. (p. 158)

Lambert and Bergin's analysis of the existing data suggests a hypothesis that is, as yet, untested by meta-analytic methods, namely, that bona fide psychotherapies are equally effective. The purpose of our meta-analysis was to test this version of the Dodo bird conjecture: When treatments intended to be therapeutic are compared, the true differences among all such treatments are zero.

Three characteristics of previous meta-analyses that have tested hypotheses related to the Dodo bird conjecture attenuate confidence in the conclusion that psychotherapies are equally efficacious: (a) Effect sizes for various types of psychotherapies were often derived from studies that did not directly compare the psychotherapies, (b) effect sizes were determined by classifying treatments into categories, and (c) the psychotherapies included were not necessarily intended to be therapeutic. Each of these problems is discussed below.

Lack of Direct Comparisons

Shadish and Sweeney (1991) demonstrated that conclusions of relative effectiveness of psychotherapies based on studies that did not directly compare the psychotherapies are confounded by differences among the studies. Consider, as did Shadish and Sweeney, the comparison of behavioral versus nonbehavioral interventions, where the effect size for each type of therapy was determined, in separate studies, by comparisons with a control group. Because the behavioral and nonbehavioral studies might differ on several other variables, such as the type of dependent variables, treatment standardization, treatment length, psychological problem treated, and so forth, the therapeutic orientation (behavioral, nonbehavioral) would be confounded with these other variables.

There are two ways to handle the confounding of treatment efficacy and other variables. First, the confounding variables can be identified and their mediating and moderating effects statistically modeled. Shadish and Sweeney (1991), for example, found that (a) setting, measurement reactivity, measurement specificity, measurement manipulability, and number of participants moderated and (b) treatment standardization, treatment implementation, and behavioral dependent variables mediated the relationship between therapeutic orientation and effect size. This modeling is subject to all of the problems inherent in any modeling of data collected passively, such as leaving out important variables, mis-specification of models, unreliability of measurements, power, and so forth. With regard to the issue of examining the appropriate mediating and moderating variables, Beutler (1991) provided a useful framework to understand how one might choose the important variables, although it becomes clear that an enormous number of variables result from considering the dimensions of therapist, patient, and process variables that have been identified as potential influences on psychotherapy outcomes.

A second way to eliminate the confounds is to consider only studies that contain direct comparisons between or among different psychotherapies. This strategy typically eliminates confounds due to any aspects of the dependent measures, the problem treated, the setting, and the length of therapy, although a few confounds remain, such as expertise of therapist and allegiance of researcher. As Shadish et al. (1993) noted, analysis of direct comparisons between therapies "have rarely been reported in past meta-analyses, and their value for controlling confounds seems to be underappreciated" (p. 998). Although there have been meta-analyses of direct comparisons of psychotherapies (e.g., Robinson, Berman, & Neimeyer, 1990; Shapiro & Shapiro, 1982), these meta-analyses have suffered from one or two of the other problems discussed in this article.

Note that direct comparisons do not rule out all confounds, although the remaining confounds likely lead to a conservative test of the Dodo bird conjecture. To understand the conservativeness of the confounds, consider the primary confound identified in the literature, researcher allegiance (e.g., Robinson et al., 1990). If allegiance to a particular treatment increases the efficacy of this treatment or decreases the efficacy of the treatment with which it is compared, then allegiance should tend to increase the size of the difference in treatment efficacy, providing evidence against the Dodo bird conjecture. (Conceivably one could have an allegiance to a less efficacious treatment, thereby reducing the effect size, although effects in this direction have not been found.) Other possible confounds, such as treatment duration, skill of therapist, and nonrandom assignment, are also likely to produce evidence against the conjecture.

Classification of Treatments Into Categories

Previous meta-analyses aimed at answering the relative efficacy question raised by the Dodo bird conjecture have compared types of treatments by classifying the various treatments studied into categories (e.g., behavioral, cognitive behavioral, and psychodynamic) and then comparing the effect sizes produced by pairwise comparisons of the classes (e.g., Robinson et al., 1990; Shapiro & Shapiro, 1982; Smith & Glass, 1977). For each pairwise comparison, one of the classes was arbitrarily designated as primary so that positive effect sizes would indicate the superiority of the designated class. For example, if behavioral treatments were designated as primary and psychodynamic ones as secondary, then the effect size for a study that compared behavioral and psychodynamic treatments would be calculated by subtracting the mean of the psychodynamic treatment from the mean of the behavioral treatment and dividing by the appropriate standard deviation—a positive effect size would indicate the superiority of the behavioral treatment. Meta-analytically, two classes were compared by using the sample effect sizes to test whether the true effect size for the pairwise comparison of the classes was zero.

Classifying treatments into classes creates problems that obviate testing the Dodo bird conjecture. The first problem is that classifications of treatments creates tests of the differences between classes of treatments whereas the Dodo bird conjecture references a general hypothesis about effects produced by treatments. In our meta-analysis, the hypothesis is that the true effect size for all comparisons between treatments is zero, as opposed to the hypothesis that the true effect size for comparisons of treatments from two selected classes is zero. A second problem, related to the first, is that classification results in pairwise tests between classes and obviates an omnibus test. Because the classification strategy and subsequent calculation of effect size requires designation of the primary class, only pairwise comparisons are possible. When such pairwise comparisons were tested, the typical result was that a few of the comparisons were statistically significant. For example, when only those classes that could be considered to represent bona fide treatments were considered, Shapiro and Shapiro (1982) found 2 significant differences out of 13; Shadish et al. (1993), 1 out of 10; and Robinson et al. (1990) 4 out of 6.[1] Without an omnibus test, it is not possible to know whether these few significant differences were due to chance. Finally, classification of treatments into classes eliminates from consideration all comparisons of treatments within classes. In the Shadish et al. (1993, Table 5) meta-analysis, over 60% of the studies compared treatments within classes. Clearly, comparisons within a class of treatments are interesting to primary researchers or they would not have conducted such comparisons; failure to include such comparisons in a meta-analysis eliminates any variation within classes from the analysis. Our meta-analysis tested the general hypothesis of relative effectiveness without classifying treatments. It could be argued that if the Dodo bird conjecture is false, it is likely that the variance of effects between classes will be larger than the variance within classes (although Shadish et al.'s analysis suggested otherwise); this possibility was tested, although in a slightly different way, in our meta-analysis.

Inclusion of Bona Fide Psychotherapies Only

The efficacy of psychotherapy has been established in clinical trials by comparing a given treatment with a waiting-list control, a placebo control, or an alternative treatment, whereas relative efficacy is established by comparing two treatments that are intended to be therapeutic (Wampold, 1997). Previous meta-analyses of direct comparisons of psychotherapies have included treatments that may not have been intended to be therapeutic. For example, 8 of 13 studies classified as *dynamic–humanistic* by Shapiro and Shapiro (1982) "contained no therapeutic elements unique to itself and could be viewed as a straw man" (p. 591). In our meta-analysis, only treatments that were intended to be therapeutic, which were labeled as *bona fide psychotherapies*, were included. Because the researchers' intent could not be assessed, we defined *bona fide psychotherapies* as those that were delivered by trained therapists and were based on psychological principles, were offered to the psychotherapy community as viable treatments (e.g., through professional books or manuals), or contained specified components. A current trend in comparative psychotherapy trials is to use alternative therapies, which are credible to the participants in the study but are not intended to be therapeutic (Wampold, 1997); consequently, we excluded alternative therapies from our definition of bona fide psychotherapies.

Our Meta-Analysis

Our meta-analysis differs from previous meta-analyses in that (a) the corpus of studies reviewed was limited to only those studies that directly compared two or more treatments, (b) treatments were not classified into general types, and (c) only bona fide psychotherapies, as defined above, were considered. Moreover, as discussed below, additional tests of the Dodo bird conjecture were created by examining the relationship of the effect sizes between psychotherapies with publication year and treatment similarity.

[1] When we were determining the number of significant comparisons, it was difficult to determine which categories contained some treatments not intended to be psychotherapeutic (e.g., alternate therapy controls). Some of the categories we discuss may contain some treatments that would not meet the operational definition of bona fide psychotherapy that we used in our meta-analysis. Nevertheless, the number of significant comparisons mentioned supports the point that the pairwise tests provide ambiguous data with regard to the relative efficacy of treatments in general.

Method

General analytic strategies used to test the Dodo bird conjecture. Eschewing classification of treatments into categories created a consequential methodological problem. For each comparison found in the primary literature, the sign to attach to the corresponding effect size was arbitrary. Simply, should the sign of the effect be positive or negative? To assign positive signs to all effects would overestimate the true effect size, although such a strategy was used to provide an upper bound to the true effect size for the difference in bona fide treatments. The alternative, to assign the signs randomly, would result in an aggregate effect size near zero. That is, if one half of the effects were randomly determined to be positive and the others negative, effects would cancel each other out. However, if there are true differences among treatments, there are relatively many comparisons whose effects are large in magnitude, producing "thick" tails of the distribution of effects with random signs (as shown in Figure 1). However, if the Dodo bird conjecture is correct, then a preponderance of the effects are near zero, as also shown in Figure 1. Our meta-analysis tested whether the distribution of effects (with random signs) was close to zero, as predicted by the Dodo bird conjecture.

Two other strategies were used to provide evidence for or against the Dodo bird conjecture. Stiles, Shapiro, and Elliott (1986), in their thoughtful analysis of the Dodo bird effect, argued that true differences in relative treatment efficacy might have been obscured by poor research methods and that as methods improve the true differences will become apparent. According to this view, it would be expected that as the years progress, there will be an improvement in research methods generally and in psychotherapy outcome research specifically (certainly, the advent of manuals would be a milepost in the development of psychotherapy research methods). These assumptions lead to the prediction that if the Dodo bird conjecture is false, then there will be a positive relation between the publication date of a study and the effect size of the contrast between bona fide therapies. However, if the true difference is zero, improved method is irrelevant and the effect sizes for comparisons of treatments will remain stably near zero.

Returning to the classification issue, we suggest another prediction. Treatments that fall into the same class are more similar to each other than treatments in different classes. Moreover, if the Dodo bird conjecture is false, then similar treatments, for example, in vivo exposure and imaginal exposure, would be expected to produce smaller differences than would treatments that are very dissimilar, such as in vivo exposure and insight-oriented treatment. Consequently, if the Dodo bird conjecture is false, then there should be a negative relation between similarity of treatment and size of the effect. Because date of publication and similarity of treatment may be confounded, the relations of effect size to date and similarity are tested simultaneously.

Selection of studies. The goal of the retrieval process was to locate the preponderance of studies that compared two or more bona fide psychotherapies. Toward this end, we decided to examine every study in the journals that typically publish such research. A preponderance of articles cited in Shapiro and Shapiro's (1982) review of comparative studies appeared in four journals: *Behavior Therapy, Behaviour Research and Therapy, Journal of Consulting and Clinical Psychology,* and *Journal of Counseling Psychology.* To this list, we added the *Archives of General Psychiatry* because it published the most comprehensive psychotherapy outcome study conducted to date (viz., the NIMH Treatment of Depression Collaborative Research Program; Elkin et al., 1989) and *Cognitive Therapy and Research* because of the proliferation of cognitive therapies in the past decade.

To be included in this meta-analysis, a study had to (a) be published in the journals listed above between (inclusively) 1970 and 1995 (a period that includes what is referred to as *Generation III research* involving clinical trials; Goldfried & Wolfe, 1996), (b) contain the necessary statistics to calculate effect sizes for the dependent measures as described below, and (c) compare two or more bona fide psychotherapies. To determine whether a treatment was a bona fide psychotherapy, we used the following selection criteria. First, the treatment must have involved a therapist with at least a master's degree and a meeting with a patient in which the therapist developed a relationship with the patient and tailored the treatment to the patient. Thus, any study that used solely tape-recorded instructions to patients or a protocol that was administered regardless of patient behavior (e.g., a progressive relaxation protocol that was not modified in any way for particular patients) was excluded. Second, the problem addressed by the treatment must have been one that would reasonably be treated by psychotherapy, although it was not required that the sample treated be classified as *clinically dysfunctional.*

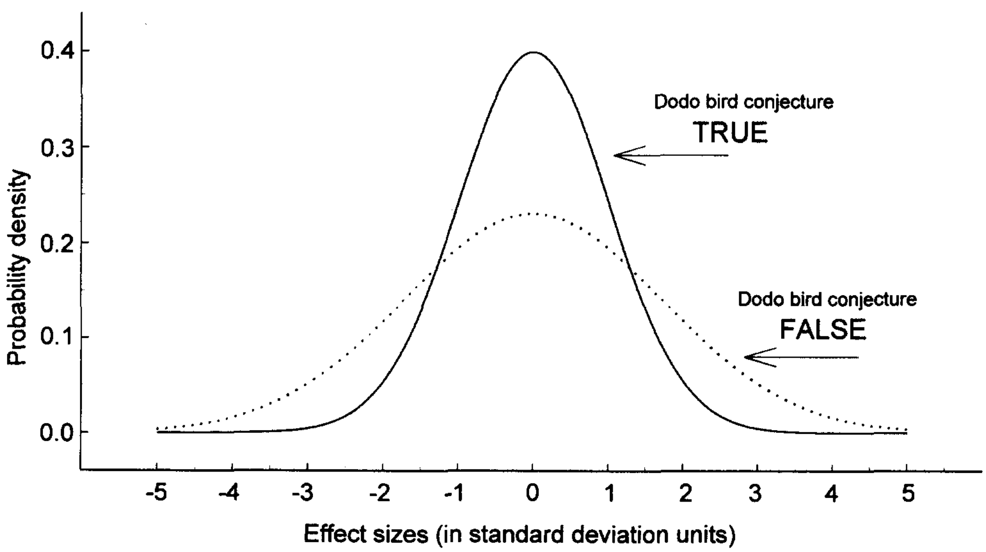

Figure 1. A distribution of effect sizes (with signs determined randomly) when the Dodo bird conjecture is true and when it is false.

For example, treatments to increase time that a participant could keep a hand submerged in cold water would be excluded because cold-water stress would not reasonably be considered a problem for which one would present to a psychotherapist. However, any treatment for depression was included whether the participants met diagnostic criteria for any depressive disorder or scored below standard cutoffs on depression scales. Finally, the treatment had to satisfy two of the following four conditions: (a) A citation was made to an established approach to psychotherapy (e.g., a reference to Rogers's, 1951, client-centered therapy), (b) a description of the therapy was contained in the article and the description contained a reference to psychological processes (e.g., operant conditioning), (c) a manual for the treatment existed and was used to guide the administration of the psychotherapy, and (d) the active ingredients of the treatment were identified and citations provided for those ingredients. Accordingly, any treatments designed to control for common or nonspecific factors, such as placebo control groups, alternative therapies, or nonspecific therapies, were excluded. Furthermore, dismantling studies were excluded because the goal of these studies was to identify active ingredients of a bona fide psychotherapy rather than to test the relative efficacy of two bona fide treatments. Finally, parametric studies that sought to vary the amount of some ingredient to determine the optimal dose were excluded.

The following procedure was used to select studies that contained two or more bona fide therapies. First, all studies that compared two or more treatments, although not rigorously defined as bona fide treatments, were retrieved by graduate research assistants. These studies were then carefully read and rated by two doctoral students in counseling psychology. If both raters indicated that two or more treatments were bona fide, using the criteria above, the study was retained, whereas if both raters indicated that the study did not contain at least two bona fide treatments, the study was rejected. In the case of a disagreement, a third rater, either another doctoral student in counseling psychology or Bruce E. Wampold, rated the study; if that rater indicated that there were more than two bona fide psychotherapies the study was retained, otherwise it was rejected. Thus, to be classified as a bona fide treatment, the treatment had to be judged by at least two of three raters as meeting the criteria discussed above.

Unit of analysis and calculation of effect size. The unit of analysis was a direct comparison of two bona fide psychotherapies. Consequently, studies that contained more than two bona fide therapies created more than one comparison. In the base analysis of the comparisons, follow-up assessments were considered distinct from assessments at termination. Thus, a study that contained three bona fide treatments (e.g., Treatments A, B, and C) and for which participants were assessed at termination and at some follow-up point resulted in six comparisons (viz., A vs. B, A vs. C, and B vs. C, each at termination and at follow-up). Finally, if the researchers tested hypotheses about the relative effectiveness of treatments on various classes of outcome measures (e.g., behavioral vs. cognitive measures) or with various subsets of participants, separate effects were coded so that the effects in the meta-analysis were sensitive to such moderating variables. For example, Gallagher-Thompson and Steffen (1994) hypothesized (and found) that a cognitive–behavioral treatment was most effective for treating caregivers with depression who had been giving care for long periods of time whereas psychodynamic therapy was most effective for treating those caregivers with depression who had been giving care for shorter periods of time; consequently, effects sizes were calculated for two groups of participants (long-term and short-term caregivers). It should be noted that including follow-up assessments and multiple treatments (i.e., more than two bona fide treatments in the same study) introduces dependencies in the data—a problem that was dealt with in our meta-analysis in various ways, as described subsequently.

The first rater extracted the following information with regard to each comparison: (a) journal name, (b) year of publication, (c) names and descriptions of the treatments, (d) number of participants in each treatment, and (e) means and standard deviations (*SD*s) for each outcome measure. This information was checked by the second rater, and disagreements were resolved by a third rater. In instances where means and *SD*s were absent, it was not possible to use inferential tests to calculate the effect size because the inferential tests either relied on data from groups other than those considered (typically a control group) or used other procedures that obviated determining the comparison of the two groups on all dependent measures. However, in some studies, the means and *SD*s were given only for those outcome measures for which some statistical significance was achieved. Ignoring the outcome measures for which data did not exist overestimated the effect size for the comparison, whereas setting the effect size for those measures to zero underestimated the effect size for the comparison. Nevertheless, to be conservative (i.e., to be biased against the Dodo bird conjecture), we conducted supplementary analyses with these studies by calculating the effect on only those measures for which means and *SD*s were provided (i.e., on the statistically significant outcome variables only).

To determine whether the overall effect size was different from zero, we found it necessary to obtain a point estimate for the true effect size for each comparison as well as an estimate for the variance of this point estimate. This was accomplished by calculating effect sizes for outcome variables within comparisons and then aggregating across outcome variables.

For each outcome measure within a comparison, the effect size g was calculated by taking the difference between the means for the two bona fide psychotherapies and dividing by the *SD* determined by pooling the *SD*s; that is,

$$g = (M_A - M_B)/s, \quad (1)$$

where M_A and M_B were the means for the variable for bona fide Treatments A and B, respectively, and s was the pooled *SD*. Although the treatment first mentioned in the article was designated as Treatment A, the sign of g was arbitrary—an issue that is addressed in our subsequent analyses. However, the sign of g was fixed so that outcomes would be consistent, regardless of the scaling of the outcome variable. That is, a positive g indicates that Treatment A was superior to Treatment B, even when low scores on the outcome indicate better psychological functioning.

Unbiased estimates of the population effect size d were calculated by correcting (approximately) for the bias in g (Hedges & Olkin, 1985):

$$d = [1 - 3/(4N - 9)]g, \quad (2)$$

where $N = n_A + n_B$, the sum of the number of participants in Treatment A and in Treatment B. The variance of d was estimated by

$$\hat{\sigma}^2(d) = [N/(n_A n_B)] + [d^2/2N]. \quad (3)$$

The estimate of the effect size for a comparison d_c was derived from a vector of the effect size d for the outcome variables and the correlation between the outcome variables, as described by Hedges and Olkin (1985). Let d_i be the vector of effect sizes and R be the correlation matrix of the outcome measures. Because the correlations of the outcome variables were rarely provided in any study, an estimate of the correlations was used. In any study, typically there are several measures of several constructs. For example, there may be two or more measures of depression and two or more measures of anxiety. In a comprehensive study of the validity of popular measures of depression and anxiety (Tanaka-Matsumi & Kameoka, 1986), the average correlation of the measures was slightly greater than .50. Because depression and anxiety are prevalent disorders (e.g., Shapiro & Shapiro, 1982, found that 56% of outcome studies targeted depression, anxiety, or both) and because it is expected that outcome measures in any study target constructs that are related, a correlation of .50 was chosen to aggregate the effect sizes within comparisons. Note that the primary effect of this aggregation is to reduce the variance vis-à-vis the variance of the estimators of the individual outcome measures (Hedges & Olkin, 1985), yielding a more precise estimate of the effect size for the comparison of two bona fide psychotherapies.

The covariance matrix Σ of d_i is $\Sigma = D_i R D_i$, where D_i is a diagonal matrix of the respective *SD*s of d_i (i.e., the square root of the variance given in Equation 3). If e is a column vector of 1s and Λ is the inverse of Σ, then the aggregate estimate of the effect size for a comparison is given by

$$d_c = [\Lambda e / e' \Lambda e] d_i, \tag{4}$$

with an estimated variance of

$$\hat{\sigma}^2(d_c) = 1/e' \Lambda e \tag{5}$$

(Hedges & Olkin, 1985, pp. 212–213). The estimates produced by Equations 4 and 5 were used in all our subsequent analyses. The strategy described above is a method to form an aggregated effect size from the several outcome measures used in a comparison of two psychotherapies.

Treatment similarity measure. Because one of the hypotheses of this study referenced the similarity of treatments, it was necessary to have a measure to operationalize similarity. A questionnaire was developed that contained one item for each comparison. The item contained the names of the two bona fide treatments, as indicated in the article, as well as one or two sentence descriptions, which were either derived from the descriptions of the treatments in the article or, in cases where the description was absent, from the sources cited in the article. Respondents were asked to rate the similarity of each pair of treatments on a 7-point scale, anchored by 7 = *very similar* and 1 = *very dissimilar.* The respondents were six academic counseling psychologists (five full professors, one associate professor), five of whom were licensed in their state of residence, from five major universities and who had a 17-year mean time since their doctoral degree and had a 58 mean number of publications in psychology-related sources. The mean response for the six academicians for each pair of treatments constituted the similarity measure. The mean rating of similarity across all items was 3.47 (*SD* = 1.17, with a range from 1.17 to 6.33), indicating that the mean ratings of the six raters spanned the possible range of similarity and that restriction of range was not problematic. The form of the intraclass correlation coefficient (ICC) designed for the mean rating of multiple judges in a fixed effects context [viz., model ICC (3,6) according to Shrout & Fleiss, 1979] was used to estimate the reliability of the ratings and equaled .87.

Tests of hypotheses. The fundamental hypothesis to test in this meta-analysis is whether the overall differential effect of bona fide psychotherapies is zero. Unfortunately, the comparison effect sizes could not be used to test directly the hypothesis that the true effect size for comparisons is zero because the sign of the effect for each comparison was arbitrary. As discussed above, two possibilities exist for assignment of signs to the effect sizes for comparisons. First, the absolute value of the comparison effect size could be used. However, if the true effect size was zero, it would be expected that about half of the sample values for the comparisons effect sizes would be negative, so this strategy provides a gross overestimate of the true effect size. Nevertheless, the aggregate of the absolute value of the comparison effect sizes is calculated and gives an upper bound of the true effect size. The estimated aggregated effect size was determined using

$$d_{|c|-\text{agg}} = \left[\sum_{i=1}^{k} \frac{d_{ci}}{\hat{\sigma}^2(d_{ci})} \right] \Big/ \left[\sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2(d_{ci})} \right], \tag{6}$$

where $d_{|c|-\text{agg}}$ is the aggregate of the set of the absolute values of the effect sizes, weighted by the inverse of the variance; d_{ci} is the estimated effect size for comparison i; and k is the number of effects aggregated (viz., number of comparisons; Hedges & Olkin, 1985, p. 111). The estimate of the variance of this aggregate is given by

$$\hat{\sigma}^2(d_{|c|-\text{agg}}) = \left(\sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2(d_{ci})} \right)^{-1} \tag{7}$$

and was used to test whether this upper bound was significantly different from zero (i.e., whether $d_{|c|-\text{agg}}$ was more than two standard errors from zero).

Another strategy would be to randomly assign a sign to the comparisons. Of course, this yields an aggregated comparison effect size near zero (calculated with Equation 6). However, if the true effect size is zero, then the effect sizes should be distributed around zero, as predicted by sampling theory for effect sizes (i.e., many effects near zero and few effects in the tails of the distribution). If the true effect size is not zero, then there should be a disproportionate number of effects in the tails of this distribution. Essentially, if the true effect size is zero, then the comparison effect sizes with random signs should be homogeneously distributed around zero. When the variances of a statistic can be estimated (which is the case here), the modified minimum chi-square test (Cramér, 1946) can be used to test for homogeneity (Hedges & Olkin, 1985). In this case,

$$Q = \sum_{i=1}^{k} \frac{(d_{ci} - d_{c-\text{agg}})^2}{\hat{\sigma}^2(d_{ci})} \tag{8}$$

and is distributed as a chi-square with $k - 1$ *df*, where k is the number of comparisons yielding effect sizes, d_{ci} is the obtained effect sizes for comparison i, and $d_{c-\text{agg}}$ is the aggregated effect size across comparisons, which in this case is assumed to be zero. Sufficiently large values of Q result in rejection of the null hypothesis of homogeneity and thus acceptance of the hypothesis that the true effect size for comparisons is nonzero.

Because there were some threats to analyses described above, modifications were made in the base data set to increase the confidence in the conclusions. The first modification involved adding effects produced by studies that reported only the means and *SD*s of outcome measures that produced statistically significant results. The second modification involved the dependencies in the data created by using follow-up assessments. In this instance, the dependency was eliminated by using only the effect produced by the longest term follow-up, if a follow-up assessment existed. The final modification was to model the dependencies amongst three bona fide treatments. Because the comparisons of Treatments A versus B and A versus C both contain Treatment A, a structural dependency was created. For each outcome measure, the estimated variances and covariances among the effects for Treatments A, B, and C, taking into account the common groups (i.e., modeling the structural dependence), were calculated according to the methods presented by Gleser and Olkin (1994, p. 346, Equations 22-12 and 22-13), then the estimated aggregate effect (and variance of aggregated effect) for each outcome measure was calculated using these variances and covariances and the effects of each comparison (see Hedges & Olkin, 1985, pp. 212–213), and finally an aggregated effect (and variance) for the comparison of the three treatments was calculated by aggregating over the outcome measures, as described above (see Equations 4 and 5).

As discussed previously, it has been contended that true differences in efficacy have been masked by relatively insensitive research methods but, as advances are made in methodology (e.g., more sensitive measures, more powerful statistical methods, better specified treatments), then true differences will be detected. If this is true, then the effect sizes should be positively related to the year in which a study is published; that is, later studies should produce larger effects.

A nonzero true effect size should also produce a negative relationship between comparison effect size and similarity. That is, if there are true differences among therapies, it should be expected that the more similar two therapies are, the smaller the differences in efficacy.

To account for any confounding between year of publication and similarity of therapies, we simultaneously regressed the comparison effects sizes onto year of publication and similarity of therapies. Because the variances of each comparison effect size had been estimated, the regression was accomplished by weighting the effect size by the inverse of the variance and adjusting the regression coefficients accordingly (Hedges & Olkin, 1985, pp. 169–175). The standard errors of the regression were corrected by dividing by the mean square error for the regression model. In this case, the relation of year and similarity to effect size were assessed by testing the regression coefficients using the adjusted standard errors.

Results

Homogeneity of effect sizes around zero. The hypothesis that the true effect size for the comparison of bona fide psychotherapies is zero was tested by randomly assigning a sign to the comparison effect sizes and calculating the modified minimum chi-square test of homogeneity. For the base data set (all comparisons for which sufficient summary statistics were presented for all outcome variables), there were 277 effects. As expected, because the signs were assigned randomly, the aggregated effect size across these effects would necessarily be near zero (indeed, $d_{|c|-agg} = .0021$). Using Equation 8 to test for homogeneity around zero, $Q = 241.18$, which was compared with a chi-square distribution with 276 *df*s, as shown in Table 1. Clearly, the value of Q was insufficient to reject the null hypothesis of homogeneity around zero ($p = .94$). When all the effects were given a positive sign, the aggregated effect size $d_{|c|-agg} = .1873$ (with a variance of .00026). Although, the aggregate of the absolute values of the comparison effect sizes clearly led to a rejection of the null hypothesis that the true effect size is zero (i.e., the aggregate is more than 2 *SE*s [standard errors] from 0), as noted previously, this estimate of the true effect is an overestimate and provides an upper bound only.

Eighteen additional effects were obtained from studies that reported only summary statistics for the statistically significant outcome measures. Overall, these studies reported statistics for 57% of the outcome variables. As shown in Table 1, adding these effects did not alter the results, as the effects remained homogeneously distributed around zero, even though including only the statistically significant variables overestimated the effect sizes for these studies.

To eliminate the dependencies induced by including termination and follow-up assessments, the data set that included only the final measurement was examined. This set included 182 effects, which were homogeneously distributed around zero (see Table 1). Finally, when the dependencies created by including three treatments in a study were modeled, the effects again were homogeneously distributed around zero. The latter data set produced the largest effect when the absolute values of the effects were aggregated (viz., .2091).

In summary, none of the databases yielded effects that vaguely approached the heterogeneity expected if there were true differences among bona fide psychotherapies. In fact, as the expected value of a chi-square distributed variate is equal to the degrees of freedom, the obtained value of Q was always very near the expected value of Q under the null hypothesis of homogeneity around zero, indicating that increasing the number of studies reviewed would not lead to a rejection of the null hypothesis (i.e., the nonsignificant results were not due to low power).

Regression analyses. Because the data set that accounted for dependencies in the data most closely met the regression assumptions (i.e., independence of observations), the regression of year of publication and similarity of treatments onto the absolute value of the effect size was conducted using the final measurement + modeling multigroup dependencies data set. The zero order correlations between year and effect size and between similarity and effect size were small (viz., −.04 and .09, respectively), although the correlation between similarity and year was larger (viz., .17, indicating that over time the similarity became greater). When the absolute value of the effect sizes was regressed onto year and similarity, using weighted least

Table 1
Tests of Homogeneity of Effects for Base Data Set and Variations

Data set	No. of effects	Q[a]	p	$d_{\|c\|-agg}$[b]	$\hat{\sigma}^2(d_{\|c\|-agg})$
Base[c]	277	241.18	.94	.1834	.00026
Base + studies without presentation of nonsignificant measures[d]	295	287.93	.59	.1917	.00025
Final measurement only[e]	182	168.06	.76	.2035	.00044
Final measurement + modeling multigroup dependencies[f]	136	137.48	.42	.2091	.00057

[a] Used to test homogeneity around zero and is compared with a chi-square statistic with $df = 1 -$ number of effects. [b] The estimated aggregated effect size over studies when all effects are given a positive sign, representing an upper bound to the true effect size for comparisons. [c] Contains all contrasts of bona fide psychotherapies for which sufficient summary statistics were presented for all dependent variables, including termination and follow-up assessments. [d] The base studies and studies that contained only summary statistics for the statistically significant outcome measures (effects are based only on reported measures). [e] Contains the effects only for the longest term follow-up if a follow-up assessment was present (termination otherwise). [f] Composed of those effects in the final measurement only data set and from which the dependency among three treatments in a single study was modeled.

squares analysis in which the weights are the inverses of the variances of the effect sizes and adjusting the standard errors (Hedges & Olkin, 1985), neither year nor similarity predicted effect size (the regression coefficient divided by the adjusted standard error were .814 and .778 for year and similarity, respectively). Effect sizes have not increased in size over the years, indicating that better research methods are not increasingly able to detect a true difference in efficacy. Similarity also was unrelated to effect size, a result consistent with the Dodo bird conjecture. Together, the regression analyses provide evidence against a claim that there are true differences among treatments, increasingly sophisticated methods are beginning to detect those differences, and comparisons of dissimilar treatments produce larger differences than do comparisons of similar treatments.

Discussion

The purpose of our meta-analysis was to test the Dodo bird conjecture, which states that when treatments intended to be therapeutic are compared, the true difference between all such treatments is zero. The results of our analysis demonstrated that the distribution of effect sizes produced by comparing two bona fide psychotherapeutic treatments was consistent with the hypothesis that the true difference is zero. Moreover, the effect sizes produced by such comparisons were not related to the similarity of the treatments compared, nor did they increase as a function of time. Finally, this study examined only direct comparisons between therapies so that the results are not confounded by differences in outcome measures. In all, the findings are entirely consistent with the Dodo bird conjecture.

Although the results of this study are consistent with an effect size of zero, criticisms could be leveled at the decision to randomly assign the sign of the effect sizes. However, the estimate of an upper bound on the effect size of comparisons between bona fide psychotherapies was determined by taking the absolute values of the effects and was found to be small (viz., less than .21). Clearly, the effect size for the comparison of bona fide psychotherapies is in the interval of .00 and .21. Table 2 summarizes the effect sizes produced by other meta-analyses (Lambert & Bergin, 1994) as well as by our meta-analysis. It is poignant to notice that the size of the effect between bona fide psychotherapies is at most about half of the effect size produced by treatments with no active psychotherapeutic ingredients (i.e., placebo vs. no treatment).

Previous Dodo bird conjectures have been called absurd (e.g., Rachman & Wilson, 1980). One of the arguments that is proposed to refute the conclusions of our meta-analysis goes along the following line: Study X found that Treatment A was superior to Treatment B, thus providing a counterexample that proves the Dodo bird conjecture false. On the contrary, one study cannot prove the Dodo bird conjecture false and even may provide evidence in favor of the conjecture. If the Dodo bird conjecture is true, then there are a few studies that produced, by chance, relatively large effect sizes; it is not possible to conclude, on the basis of a single or a few studies, that the Dodo bird conjecture is false. Rather, the conjecture must be tested on the corpus of studies, as was accomplished in our meta-analysis. Moreover, the conclusion that Treatment A was superior to Treatment B was likely made on the basis of the statistical significance achieved by a few of many outcome measures. Given the assumption that researchers choose outcome measures that are germane to the psychological functioning of the patients involved in the study, it is the effect of the treatment on the set of outcome measures that is important. Our meta-analysis accounted for this by aggregating the effect sizes across outcome measures. Focusing on a few of many outcome measures to establish superiority causes fishing and error rate problems (Cook & Campbell, 1979) and distracts the researcher from examining the set of outcome measures, which might have produced a negligible effect size. In any event, a single study, no matter how large the aggregated effect size, cannot disprove the Dodo bird conjecture.

Table 2
Summary of Effect Sizes Produced by Meta-Analyses of Psychotherapy Outcomes

Comparison	Size of effect
Psychotherapy vs. no treatment	.82
Psychotherapy vs. placebo	.48
Placebo vs. no treatment	.42
Differences between bona fide psychotherapies	.00 < ES < .21

Note. Effect sizes (ESs) for the first three comparisons were derived by Lambert and Bergin (1994) from extant meta-analyses.

There are a number of limitations of this study that limit the scope of the conclusions that can be made. One issue is that the entire corpus of comparison studies was not retrieved. Because key words typically refer to substantive findings of a study rather than to methodological features, retrieval of comparative studies using databases (e.g., *PsycLIT*) was impractical. The process of collecting studies from selected journals omitted articles published in other journals and unpublished studies. However, because statistically significant differences between bona fide psychotherapies are relatively rare and are interesting to the psychotherapy community, it is unlikely that studies containing such differences would be unpublished, unless the study was flawed. Note that excluding unpublished studies that did not contain differences between therapies would produce a bias against the Dodo bird conjecture.

The sampling of treatments and disorders presents another issue that needs consideration. That there are about 250 types of therapy and 300 disorders (Goldfried & Wolfe, 1996) clearly indicates that the comparisons reviewed for this meta-analysis were not sampled from a Types of Therapy × Types of Disorder matrix. A perusal of studies reviewed indicates an overrepresentation of behavioral and cognitive–behavior treatments. Moreover, some therapies are specific to disorders (e.g., exposure treatments for phobias), whereas others are more appropriate for a wide variety of disorders (e.g., cognitive therapies). Conversely, some disorders are amenable to many treatments (e.g., depression), whereas others may not be (e.g., obsessive–compulsive disorder). Consequently, it would be unwarranted to conclude from this study that all therapies are equally effective with all disorders.

Moreover, it should be recognized that the psychotherapies studied were those bona fide psychotherapies selected by psychotherapy researchers. At least two types of therapies practiced in the "real world" were not included in the studies analyzed. First, some psychotherapies would not meet the operational definition used in this study (e.g., they are not based on psychological principles). Second, other practiced therapies would meet the criteria for being classified as bona fide but were not studied by psychotherapy researchers. Consequently, the results of this meta-analysis should not be construed to support the conclusion that all practiced psychotherapies are equally efficacious or are as efficacious as the ones we reviewed.

The aggregation strategies used in this meta-analysis create a few issues. This meta-analysis accounted for interactions between treatments and person variables and differential effects on various subsets of dependent measures, provided they were hypothesized and tested by the primary researchers. However, it is likely that such interactive effects are present in the primary studies but not hypothesized nor measured (Lyons & Howard, 1991). As is the case with the primary studies, the results reference average effects; but it is not appropriate to conclude that every treatment is equally effective with every patient. The results of this meta-analysis suggest that the efficacy of the treatments are comparable, not that the treatments are interchangeable. Another problem related to aggregation is that researchers may have included dependent variables that were either unimportant or not fully intended to be sensitive to the treatment without making these intentions explicit. It is recommended that researchers in the future clearly identify centrally important outcome variables.

It is unrealistic to believe that clinical trials will ever be conducted to cover even a portion of the possible Treatment × Disorder cells (Goldfried & Wolfe, 1996). Consequently, an omnibus hypothesis about all treatments with all disorders will never be tested. The results of this meta-analysis are robust to the threat that researchers compared only treatments that were similar because the mean similarity rating was about at the midpoint of the similarity scale, the range spanned the scale, and—most important—the similarity ratings were unrelated to effect size. Essentially, the results of this study have shown that the treatments that have interested researchers do not produce even weak evidence of differential effectiveness and provide the most global conclusion that can be made, given extant research examining direct comparisons of psychotherapeutic treatments.

In spite of the limitations, the uniform effectiveness of the treatments reviewed have profound implications for research and practice. An identification of a set of treatments that are empirically validated has been linked to the survival of applied psychology, given advances in biological psychiatry (Task Force on Promotion, 1995). The goal of the empirical validation movement is to identify a small set of treatments that satisfy criteria, which are based on the assumption that the unique ingredients of the treatment are responsible for the efficacy of the treatment (Wampold, 1997). Unfortunately, the empirical validation strategy weakens support for psychotherapy as a mental health treatment rather than strengthens it. Klein (1996), an advocate of psychopharmacological treatments, summed up the issue succinctly: "The bottom line is that if the Food and Drug Administration (FDA) was responsible for the evaluation of psychotherapy, then no current psychotherapy would be approvable, whereas particular medications are clearly approvable" (p. 84). The basis of this bold statement is that the FDA requires that the efficacy of active ingredients of any medication be established. Klein argued cogently that cognitive–behavioral therapy (CBT) for depression is contra-indicated because (a) CBT has not been shown to be more effective than placebo conditions (Robinson et al., 1990), (b) various psychotherapies do not differ in terms of efficacy, and (c) CBT did not appear to be effective with those patients for whom it was indicated, denigrating the importance of the active ingredients in CBT (see Jacobson et al., 1996, for results that were unable to validate the active ingredients of CBT). Klein (1996) concluded, in an attempt to damn psychotherapy, that

> [the results of the NIMH study and other studies] are inexplicable on the basis of the therapeutic action theories propounded by the creators of IPT [interpersonal psychotherapy] and CBT. However they are entirely compatible with the hypothesis (championed by Jerome Frank; see Frank & Frank, 1991) that psychotherapies are not doing anything specific; rather, they are nonspecifically beneficial to the final common pathway of demoralization, to the degree that they are effective at all. (p. 82)

Klein's (1996) criticism is painful only if one buys into the necessity of validating psychotherapy based on the active ingredients. If one gives up the belief that psychotherapy treatments are analogous to medications and places faith in the scientific evidence that psychotherapy in general is extremely efficacious (Lambert & Bergin, 1994) but that relative differences are minimal, research in psychotherapy would differ considerably from the present focus on clinical trials. Why is it that researchers persist in attempts to find treatment differences, when they know that these effects are small in comparison to other effects, such as therapists effects (Crits-Christoph et al., 1991; Crits-Christoph & Mintz, 1991) or effects of treatment versus no-treatment comparisons (Lambert & Bergin, 1994)?

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Alden, L. E. (1988). Behavioral self-management controlled drinking strategies in a context of secondary prevention. *Journal of Consulting and Clinical Psychology, 56,* 280–286.

*Andrews, W. R. (1971). Behavioral and client-centered counseling of high school underachievers. *Journal of Counseling Psychology, 18,* 93–96.

*Arean, P. A., Perri, M. G., Nezu, A. M., Schein, F. C., Christopher, F., & Joseph, T. X. (1993). Comparative effectiveness of social problem-solving therapy and reminiscence therapy as treatments for depression in older adults. *Journal of Consulting and Clinical Psychology, 61,* 1003–1010.

*Arnow, B. A., Taylor, C. B., Agras, W. S., & Telch, M. J. (1985). Enhancing agoraphobia treatment outcomes by changing couple communication patterns. *Behavior Therapy, 16,* 452–467.

*Barlow, D. H., Craske, M. G., Cerny, J. A., & Klosko J. S. (1989). Behavioral treatment of panic disorder. *Behavior Therapy, 20,* 261–282.

*Barlow, D. H., Rapee, R. M., & Brown, T. A. (1992). Behavioral treatment of generalized anxiety disorder. *Behavior Therapy, 23,* 551–570.

*Beach, S. R. H., & O'Leary, K. D. (1992). Treating depression in the context of marital discord: Outcome and predictors of response of marital therapy versus cognitive therapy. *Behavior Therapy, 23,* 507–528.

*Beck, J. G., Stanley, M. A., Baldwin, L. E., Deagle, E. A., III, & Averill P. A. (1994). Comparison of cognitive therapy and relaxation training for panic disorder. *Journal of Consulting and Clinical Psychology, 62,* 818–826.

Beutler, L. E. (1991). Have all won and must all have prizes? Revisiting Luborsky et al.'s verdict. *Journal of Consulting and Clinical Psychology, 59,* 226–232.

*Borkovec, T. D., & Costello, E. (1993). Efficacy of applied relaxation and cognitive–behavioral therapy in the treatment of generalized anxiety disorder. *Journal of Consulting and Clinical Psychology, 61,* 611–619.

*Borkovec, T. D., & Mathews, A. M. (1988). Treatment of nonphobic anxiety disorders: A comparison of nondirective, cognitive, and coping desensitization therapy. *Journal of Consulting and Clinical Psychology, 56,* 877–884.

*Bramston, P., & Spence, S. H. (1985). Behavioural versus cognitive social-skills training with intellectually-handicapped adults. *Behaviour Research and Therapy, 23,* 239–246.

*Butler, G., Fennell, M., Robson, P., & Gelder, M. (1991). Comparison of behavior therapy and cognitive behavior therapy in the treatment of generalized anxiety disorder. *Journal of Consulting and Clinical Psychology, 59,* 167–175.

*Carmody, T. P. (1978). Rational–emotive, self-instructional, and behavioral assertion training: Facilitating maintenance. *Cognitive Therapy and Research, 2,* 241–253.

Carroll, L. (1962). *Alice's adventures in wonderland.* Harmondsworth, Middlesex, England: Penguin Books. (Original work published 1865)

*Chang-Liang, R., & Denney, D. R. (1976). Applied relaxation as training in self-control. *Journal of Counseling Psychology, 23,* 183–189.

*Cianni, M., & Horan, J. J. (1990). An attempt to establish the experimental construct validity of cognitive and behavioral approaches to assertiveness training. *Journal of Counseling Psychology, 37,* 243–247.

*Clarke, K. M., & Greenberg, L. S. (1986). Differential effects of the gestalt two-chair intervention and problem solving in resolving decisional conflict. *Journal of Counseling Psychology, 33,* 11–15.

*Collins, R. L., Rothblum, E. D., & Wilson, G. T. (1986). The comparative efficacy of cognitive and behavioral approaches to the treatment of obesity. *Cognitive Therapy and Research, 10,* 299–318.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings.* Chicago: Rand McNally.

*Cooper, P. J., & Steere, J. (1995). A comparison of two psychological treatments for bulimia nervosa: Implications for models of maintenance. *Behaviour Research and Therapy, 33,* 875–885.

Cramér, H. (1946). *Mathematical methods of statistics.* Princeton, NJ: Princeton University Press.

*Craske, M. G., Brown, T. A., & Barlow, D. H. (1991). Behavioral treatment of panic disorder: A two-year follow-up. *Behavior Therapy, 22,* 289–384.

*Craske, M. G., Street, L., & Barlow, D. H. (1989). Instructions to focus upon or distract from internal cues during exposure treatment of agoraphobic avoidance. *Behaviour Research and Therapy, 27,* 663–672.

Crits-Christoph, P., Baranackie, K., Kurcias, J. S., Carroll, K., Luborsky, L., McLellan, T., Woody, G., Thompson, L., Gallagier, D., & Zitrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 1,* 81–91.

Crits-Christoph, P., & Mintz, J. (1991). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *Journal of Consulting and Clinical Psychology, 59,* 20–26.

*Curran, J. P., Gilbert, F. S., & Little, L. M. (1976). A comparison between behavior replication training and sensitivity training approaches to heterosexual dating anxiety. *Journal of Counseling Psychology, 23,* 190–196.

*Deffenbacher, J. L. (1988). Cognitive-relaxation and social skills treatment of anger: A year later. *Journal of Counseling Psychology, 35,* 234–236.

*Deffenbacher, J. L., & Hahnloser, R. M. (1981). Cognitive and relaxation coping skills in stress inoculation. *Cognitive Therapy and Research, 5,* 211–215.

*Deffenbacher, J. L., Mathis, H., & Michaels, A. C. (1979). Two self-control procedures in the reduction of targeted and nontargeted anxieties. *Journal of Counseling Psychology, 26,* 120–127.

*Deffenbacher, J. L., & Michaels, A. C. (1980). Two self-control procedures in the reduction of targeted and nontargeted anxieties—A year later. *Journal of Counseling Psychology, 27,* 9–15.

*Deffenbacher, J. L., Oetting, E. R., Huff, M. E., & Thwaites, G. A. (1995). Fifteen-month follow-up of social skills and cognitive-relaxation approaches to general anger reduction. *Journal of Counseling Psychology, 42,* 400–405.

*Deffenbacher, J. L., & Shelton, J. L. (1978). Comparison of anxiety management training and desensitization in reducing test and other anxieties. *Journal of Counseling Psychology, 25,* 277–282.

*Deffenbacher, J. L., Story, D. A., Brandon, A. D., Hogg, J. A., & Hazaleus, S. L. (1988). Cognitive and cognitive-relaxation treatments of anger. *Cognitive Therapy and Research, 12,* 167–184.

*Deffenbacher, J. L., Story, D. A., Stark, R. S., Hogg, J. A., & Brandon, A. D. (1987). Cognitive-relaxation and social skills interventions in the treatment of general anger. *Journal of Counseling Psychology, 34,* 171–176.

*Deffenbacher, J. L., Thwaites, G. A., Wallace, T. L., & Oetting, E. R. (1994). Social skills and cognitive-relaxation approaches to general anger reduction. *Journal of Counseling Psychology, 41,* 386–396.

*Dendato, K. M., & Diener, D. (1986). Effectiveness of cognitive relaxation therapy and study-skills training in reducing self-reported anxiety and improving the academic performance of test-anxious students. *Journal of Counseling Psychology, 33,* 131–135.

*Denney, D. R., & Sullivan, B. J. (1976). Desensitization and modeling treatments of spider fear using two types of scenes. *Journal of Consulting and Clinical Psychology, 44,* 573–579.

*De Ruiter, C., Rijken, H., Garssen, B., & Kraaimaat, F. (1989). Breathing retraining, exposure and a combination of both, in the treatment of panic disorder with agoraphobia. *Behaviour Research and Therapy, 27,* 647–655.

*D'Zurilla, T. J., Wilson, G. T., & Nelson, R. (1973). A preliminary study of the effectiveness of graduated prolonged exposure in the treatment of irrational fear. *Behavior Therapy, 4,* 672–685.

Elkin, I., Gibbons, R. D., Shea, M. T., & Shaw, B. F. (1996). Science is not a trial (but it can sometimes be a tribulation). *Journal of Consulting and Clinical Psychology, 64,* 92–103.

*Elkin, I., Shea, T., Watkins, J. T., Imber, S. D., Sotsky, S. M., & Collins, J. F. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Archives of General Psychiatry, 46,* 971–982.

*Emmelkamp, P. M. G. (1974). Self-observation versus flooding in the treatment of agoraphobia. *Behaviour Research and Therapy, 12,* 229–237.

*Emmelkamp, P. M. G., & Beens, H. (1991). Cognitive therapy with obsessive–compulsive disorder: A comparative evaluation. *Behaviour Research and Therapy, 29,* 293–300.

*Emmelkamp, P. M. G., Visser, S., & Hoekstra, R. J. (1988). Cognitive therapy vs exposure in vivo in the treatment of obsessive–compulsives. *Cognitive Therapy and Research, 12,* 103–114.

*Emmelkamp, P. M. G., & Wessels, H. (1975). Flooding in imagination vs flooding in vivo: A comparison with agoraphobics. *Behaviour Research and Therapy, 13,* 7–15.

*Fairburn, C. G., Kirk, J., O'Conner, M., & Cooper, P. J. (1986). A comparison of two psychological treatments for bulimia nervosa. *Behaviour Research and Therapy, 24,* 629–643.

*Foa, E. B., Rothbaum, B. O., Riggs, D. S., & Murdock, T. B. (1991). Treatment of posttraumatic stress disorder in rape victims: A comparison between cognitive–behavioral procedures and counseling. *Journal of Consulting and Clinical Psychology, 59,* 715–723.

*Foa, E. B., Steketee, G., & Grayson, J. B. (1985). Imaginal and in vivo exposure: A comparison with obsessive–compulsive checkers. *Behavior Therapy, 16,* 292–302.

*Foa, E. B., Steketee, G., Grayson, J. B., Turner, R. M., & Latimer, P. R. (1984). Deliberate exposure and blocking of obsessive–compulsive rituals: Immediate and long term effects. *Behavior Therapy, 15,* 450–472.

*Freeling, N. W., & Shemberg, K. M. (1970). The alleviation of test anxiety by systematic desensitization. *Behavior Research and Therapy, 8,* 293–299.

*Fremouw, W. J., & Zitter, R. E. (1978). A comparison of skills training and cognitive restructuring–relaxation for the treatment of speech anxiety. *Behavior Therapy, 9,* 248–259.

*Gallager-Thompson, D., & Steffen, A. M. (1994). Comparative effects of cognitive–behavioral and brief psychodynamic psychotherapies for depressed family caregivers. *Journal of Consulting and Clinical Psychology, 62,* 543–549.

*Gauthier, J., Pellerin, D., & Renaud, P. (1983). The enhancement of self-esteem: A comparison of two cognitive strategies. *Cognitive Therapy and Research, 7,* 389–398.

Gleser, L. J., & Olkin, I. (1994). Stochastically dependent effect sizes. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 339–355). New York: Russell Sage Foundation.

Goldfried, M. R., & Wolfe, B. E. (1996). Psychotherapy practice and research: Repairing a strained alliance. *American Psychologist, 51,* 1007–1016.

*Goldman, A., & Greenberg, L. (1992). Comparison of integrated systemic and emotionally focused approaches to couples therapy. *Journal of Consulting and Clinical Psychology, 60,* 962–969.

*Gormally, J., Varvil-Weld, D., Raphael, R., & Sipps, G. (1981). Treatment of socially anxious college men using cognitive counseling and skills training. *Journal of Counseling Psychology, 28,* 147–157.

*Hazaleus, S. L., & Deffenbacher, J. L. (1986). Relaxation and cognitive treatments of anger. *Journal of Consulting and Clinical Psychology, 54,* 222–226.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

*Hodgson, J. W. (1981). Cognitive versus behavioral–interpersonal approaches to the group treatment of depressed college students. *Journal of Counseling Psychology, 28,* 243–249.

*Hoffart, A. (1995). A comparison of cognitive and guided mastery therapy of agoraphobia. *Behaviour Research and Therapy, 33,* 423–434.

*Hogg. J. A., & Deffenbacher, J. L. (1988). A comparison of cognitive and interpersonal-process group therapies in the treatment of depression among college students. *Journal of Counseling Psychology, 35,* 304–310.

*Holmes, D. P., & Horan, J. J. (1976). Anger induction in assertion training. *Journal of Counseling Psychology, 23,* 108–111.

*Hutchings, D. F., Denney, D. R., Basgall, J., & Houston B. K. (1980). Anxiety management and applied relaxation in reducing general anxiety. *Behaviour Research and Therapy, 18,* 181–190.

Jacobson, N. (1991). Behavioral versus insight-oriented marital therapy: Labels can be misleading. *Journal of Consulting and Clinical Psychology, 59,* 142–145.

*Jacobson, N. S., Dobson, K., Fruzzetti, A. E., Schmaling, K. B., & Salusky, S. (1991). Marital therapy as a treatment for depression. *Journal of Consulting and Clinical Psychology, 59,* 547–557.

Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., Gortner, E., & Prince, S. E. (1996). A component analysis of cognitive–behavioral treatment for depression. *Journal of Consulting and Clinical Psychology, 64,* 295–304.

Jacobson, N. S., & Hollon, S. D. (1996a). Cognitive–behavior therapy versus pharmacotherapy: Now that the jury's returned its verdict, it's time to present the rest of the evidence. *Journal of Consulting and Clinical Psychology, 64,* 74–80.

Jacobson, N. S., & Hollon, S. D. (1996b). Prospects for future comparisons between drugs and psychotherapy: Lessons from the CBT-versus-pharmacotherapy exchange. *Journal of Consulting and Clinical Psychology, 64,* 104–108.

*Jenni, M. A., & Wollersheim, J. P. (1979). Cognitive therapy, stress management training, and the Type A behavior pattern. *Cognitive Therapy and Research, 3,* 61–73.

*Jerremalm, A., Jansson, L., & Ost, L. (1986a). Cognitive and physiological reactivity and the effects of different behavioral methods in the treatment of social phobia. *Behaviour Research and Therapy, 24,* 171–180.

*Jerremalm, A., Jansson, L., & Ost, L. (1986b). Individual response patterns and the effects of different behavioral methods in the treatment of dental phobia. *Behaviour Research and Therapy, 24,* 587–596.

*Kanter, N. J., & Goldfried, M. R. (1979). Relative effectiveness of rational restructuring and self-control desensitization in the reduction of interpersonal anxiety. *Behavior Therapy, 10,* 472–490.

*Kazdin, A. E., Esveldt-Dawson, K., French, N. H., & Unis, A. S. (1987). Problem-solving skills training and relationship therapy in the treatment of antisocial child behavior. *Journal of Consulting and Clinical Psychology, 55,* 76–85.

*Kelly, K. R., & Stone, G. L. (1987). Effects of three psychological treatments and self-monitoring on the reduction of Type A behavior. *Journal of Counseling Psychology, 34,* 46–54.

*Kendrick, M. J., Craig, K. D., Lawson, D. M., & Davidson, P. O. (1982). Cognitive and behavioral therapy for musical-performance anxiety. *Journal of Consulting and Clinical Psychology, 50,* 353–362.

*Kipper, D. A., & Giladi, D. (1978). Effectiveness of structured psychodrama and systematic desensitization in reading test anxiety. *Journal of Counseling Psychology, 25,* 499–505.

*Kirkland, K., & Hollandsworth, J. G. (1980). Effective test taking: Skills-acquisition versus anxiety-reduction techniques. *Journal of Clinical and Consulting Psychology, 48,* 431–439.

*Kirsch, I., & Henry, D. (1979). Self-desensitization and meditation in the reduction of public speaking anxiety. *Journal of Clinical and Consulting Psychology, 47,* 536–541.

Klein, D. F. (1996). Preventing hung juries about therapy studies. *Journal of Consulting and Clinical Psychology, 64,* 81–87.

Lambert, M. J., & Bergin, A. E. (1994) . The effectiveness of psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 143–189). New York: Wiley.

*Leal, L. L., Baxter, E. G., Martin, J., & Marx, R. W. (1981). Cognitive modification and systematic desensitization with test anxious high school students. *Journal of Counseling Psychology, 28,* 525–528.

*Lent, R. W., & Russell, R. K. (1978). Treatment of anxiety by cue-controlled desensitization and study-skills training. *Journal of Counseling Psychology, 25,* 217–224.

*Lipsky, M. J., Kassinove, H., & Miller, N. J. (1980). Effects of rational–emotive therapy, rational role reversal, and rational–emotive imagery on the emotional adjustment of community mental health center patients. *Journal of Consulting and Clinical Psychology, 48,* 366–374.

*Lomont, J. F., & Sherman, L. J. (1971). Group systematic desensitization and group insight therapies for test anxiety. *Behavior Therapy, 2,* 511–518.

Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies: Is it true that "everyone has won and all must have prizes"? *Archives of General Psychiatry, 32,* 995–1008.

Lyons, J. S., & Howard, K. I. (1991). Main effects analysis in clinical research: Statistical guidelines for disaggregating treatment groups. *Journal of Consulting and Clinical Psychology, 59,* 745–748.

*Mavissakalian, M., Michelson, L., Greenwald, D., Kornblith, S., & Greenwald, M. (1983). Cognitive–behavioral treatment of agoraphobia: Paradoxical intention vs self-statement training. *Behaviour Research and Therapy, 21,* 75–86.

*Meichenbaum, D. H. (1972). Cognitive modification of test anxious college students. *Journal of Consulting and Clinical Psychology, 39,* 370–380.

*Melnick, J., & Russell, R. W. (1976). Hypnosis versus systematic desensatization in the treatment of test anxiety. *Journal of Counseling Psychology, 23,* 291–295.

*Michelson, L., Mavissakalien, M., & Marchione, K. (1988). Cognitive, behavioral, and psychophysiological treatments of agoraphobia: A comparative outcome investigation. *Behavior Therapy, 19,* 97–120.

*Moon, J. R., & Eisler, R. M. (1983). Anger control: An experimental comparison of three behavioral treatments. *Behavior Therapy, 14,* 493–505.

*Morin, C. M., & Azrin, N. H. (1988). Behavioral and cognitive treatments of geriatric insomnia. *Journal of Consulting and Clinical Psychology, 56,* 748–753.

*Murphy, A. I., Lehrer, P. M., & Jurish, S. (1990). Cognitive coping skills training and relaxation training as treatments for tension headaches. *Behavior Therapy, 21,* 89–98.

*Newman, A., & Brand, E. (1980). Coping responses training versus in vivo desensitization in fear reduction. *Cognitive Therapy and Research, 4,* 397–407.

*Nicholas, M. K., Wilson, P. H., & Goyen, J. (1991). Operant–behavioural and cognitive–behavioural treatment for chronic low back pain. *Behaviour Research and Therapy, 29,* 225–238.

*O'Farrell, T. J., Cutter, H. S. G., Choquette, K. A., Floyd, F. J., & Bayog, R. D. (1992). Behavioral martial therapy for male alcoholics: Marital and drinking adjustment during the two years following treatment. *Behavior Therapy, 23,* 529–549.

*O'Farrell, T. J., Cutter, H. S. G., & Frank, F. J. (1985). Evaluating behavioral marital therapy for male alcoholics: Effects on marital adjustment and communication before and after treatment. *Behavior Therapy, 16,* 147–167.

Omer, H., & Dar, R. (1992). Changing trends in three decades of psychotherapy research: The flight from theory into pragmatics. *Journal of Consulting and Clinical Psychology, 60,* 88–93.

*Ost, L. G., Jerremalm, A., & Jansson, L. (1984). Individual response patterns and the effects of different behavioral methods in the treatment of agoraphobia. *Behaviour Research and Therapy, 22,* 697–707.

*Ost, L., Johansson, J., & Jerremalm, A. (1982). Individual response patterns and the effects of different behavioral methods in the treatment of claustrophobia. *Behaviour Research and Therapy, 20,* 445–460.

*Osterhouse, R. A. (1972). Desensitization and study-skills training as treatment for two types of test-anxious students. *Journal of Counseling Psychology, 19,* 301–307.

*Padfield, M. (1976). The comparative effects of two counseling approaches on the intensity of depression among rural women of low socioeconomic status. *Journal of Counseling Psychology, 23,* 209–214.

Rachman, S. J., & Wilson, G. T. (1980). *The effects of psychological therapy* (2nd ed.). New York: Pergamon Press.

*Rehm, L. P., Kaslow, N. J., & Rabin, A. S. (1987). Cognitive and behavioral targets in a self-control therapy program for depression. *Journal of Consulting and Clinical Psychology, 55,* 60–67.

*Resick, P. A., Jordan, C. G., Girelli, S. A., Hutter, C. K., & Marhoefer-Dvorak, S. (1988). A comparative outcome study of behavioral group therapy for sexual assault victims. *Behavior Therapy, 19,* 385–401.

*Reynolds, W. M., & Coats, K. I. (1986). A comparison of cognitive–behavioral therapy and relaxation training for the treatment of depression in adolescents. *Journal of Consulting and Clinical Psychology, 54,* 653–660.

Robinson, L. A., Berman, J. S., & Neimeyer, R. A. (1990). Psychotherapy for the treatment of depression: A comprehensive review of controlled outcome research. *Psychological Bulletin, 108,* 30–49.

*Rodriguez, M., & Blocher, D. (1988). A comparison of two approaches to enhancing career maturity in Puerto Rican college women. *Journal of Counseling Psychology, 35,* 275–280.

Rogers, C. R. (1951). *Client-centered therapy: Its current practice, implications, & theory.* Boston: Houghton Mifflin.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods in psychotherapy. *American Journal of Orthopsychiatry, 6,* 412–415.

*Russell, R. K., Miller, D. E., & June, L. N. (1975). A comparison between group systematic desensitization and cue-controlled relaxation in the treatment of test anxiety. *Behavior Therapy, 6,* 172–177.

*Russell, R. K., & Wise, F. (1976). Treatment of speech anxiety by cue-controlled relaxation and desensitization with professional and paraprofessional counselors. *Journal of Counseling Psychology, 23,* 583–586.

*Russell, R. K., Wise, F., & Stratoudakis, J. P. (1976). Treatment of anxiety by cue-controlled relaxation and systematic desensitization. *Journal of Counseling Psychology, 23,* 563–566.

*Safan, J. D., Alden, L. E., & Davidson, P. O. (1980). Client anxiety level as a moderator variable in assertion training. *Cognitive Therapy and Research, 4,* 189–200.

Shadish, W. R., Montgomery, L. M., Wilson, P., Wilson, M. R., Bright, I., & Okwumabua, T. (1993). Effects of family and marital psychotherapies: A meta-analysis. *Journal of Consulting and Clinical Psychology, 61,* 992–1002.

Shadish, W. R., Jr., & Sweeney, R. B. (1991). Mediators and moderators in meta-analysis: There's a reason we don't let Dodo birds tell us which psychotherapies should have prizes. *Journal of Consulting and Clinical Psychology, 59,* 883–893.

*Shapiro, D. A., Barkham, M., Rees, A., Hardy, G. E., & Reynolds, S. (1994). Effects of treatment duration and severity of depression on the effectiveness of cognitive–behavioral and psychodynamic–interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology, 62,* 522–534.

Shapiro, D. A., & Shapiro, D. (1982). Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin, 92,* 581–604.

*Shear, M. K., Pilkonis, P. A., Cloitre, M., & Leon, A. C. (1994). Cognitive behavioral treatment compared with nonprescriptive treatment of panic disorder. *Archives of General Psychiatry, 51,* 395–401.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32,* 752–760.

Snyder, D. K., Wills, R. M., & Grady-Fletcher, A. (1991). Long-term effectiveness of behavioral versus insight-oriented marital therapy: A

4-year follow-up study. *Journal of Consulting and Clinical Psychology, 59,* 138–141.
*Steuer, J. L., Mintz, J., Hammen, C. L., Hill, M. A., Jarvik, L. F., & McCarley, T. (1984). Cognitive–behavioral and psychodynamic group psychotherapy in treatment of geriatric depression. *Journal of Consulting and Clinical Psychology, 52,* 180–189.
Stiles, W. B., Shapiro, D. A., & Elliott, R. (1986). "Are all psychotherapies equivalent?" *American Psychologist, 41,* 165–180.
*Sweeny, G. A., & Horan, J. J. (1982). Separate and combined effects of cue-controlled relaxation and cognitive restructuring in the treatment of musical performance anxiety. *Journal of Counseling Psychology, 29,* 486–497.
Tanaka-Matsumi, J., & Kameoka, V. A. (1986). Reliabilities and concurrent validities of popular self-report measures of depression, anxiety, and social desirability. *Journal of Consulting and Clinical Psychology, 54,* 328–333.
Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically-validated psychological treatment: Report and recommendations. *The Clinical Psychologist, 48,* 2–23.
*Taylor, F. G., & Marshall, W. L. (1977). Experimental analysis of a cognitive–behavioral therapy for depression. *Cognitive Therapy and Research, 1,* 59–72.
*Thackwray, D. E., Smith, M. C., Bodfish, J. W., & Myers, A. W. (1993). A comparison of behavioral and cognitive–behavioral interventions for bulimia nervosa. *Journal of Consulting and Clinical Psychology, 61,* 639–645.
*Thompson, L. W., Gallagher, D., & Breckenridge, J. S. (1987). Comparative effectiveness of psychotherapies for depressed elders. *Journal of Consulting and Clinical Psychology, 55,* 385–390.
*Thurman, C. W. (1985). Effectiveness of cognitive–behavioral treatment in reducing Type A behavior among university faculty. *Journal of Counseling Psychology, 32,* 74–83.
*Turner, J. A., & Clancy, S. (1988). Comparison of operant behavioral and cognitive–behavioral group treatment for chronic low back pain. *Journal of Consulting and Clinical Psychology, 56,* 261–266.
*Turner, R. M., & Acher, L. M. (1979). Controlled comparison of progressive relaxation, stimulus control, and paradoxical intention therapies for insomnia. *Journal of Consulting and Clinical Psychology, 47,* 500–508.
*Van Oppen, P., De Hann, E., Van Balkom, A. J. L. M., Spinhoven, P., Hoogduin, K., & Van Dyck, R. (1995). Cognitive therapy and exposure in vivo in the treatment of obsessive compulsive disorder. *Behaviour Research and Therapy, 33,* 379–390.
Wampold, B. E. (1997). Methodological problems in identifying efficacious psychotherapies. *Psychotherapy Research, 7,* 21–43.
*Warren, R., McLellarn R. W., & Ponzoha, C. (1988). Rational–emotive therapy vs general cognitive–behavioral therapy in the treatment of low self-esteem and related emotional disturbances. *Cognitive Therapy and Research, 12,* 21–37.
*Watkins, J. T., Leber, W. R., Imber, S. D., Collins, J. F., Elkin, I., & Pilkonis, P. A. (1993). Temporal course of change of depression. *Journal of Consulting and Clinical Psychology, 61,* 858–864.
*Weissberg, M. (1977). A comparison of direct and vicarious treatments of speech anxiety: Desensitization, desensitization with coping imagery, and cognitive modification. *Behavior Therapy, 8,* 606–620.
*Wilfey, D. E., Agras, W. S., Telch, C. F., Rossiter, E. M., Schneider, J. A., & Cole, A. G. (1993). Group cognitive–behavioral therapy and group interpersonal psychotherapy for the nonpurging bulimic individual: A controlled comparison. *Journal of Consulting and Clinical Psychology, 61,* 296–305.
*Williams, S. L., & Zane, G. (1989). Guided mastery and stimulus exposure treatments for severe performance anxiety in agoraphobics. *Behaviour Research and Therapy, 27,* 237–245.
*Wilson, G. T., Rossiter, E., Kleifield, E. I., & Lindholm, L. (1986). Cognitive–behavioral treatment of bulimia nervosa: A controlled evaluation. *Behaviour Research and Therapy, 24,* 277–288.
*Wilson, P. H., Goldin, J. C., & Charbonneau-Powis, M. (1983). Comparative efficacy of behavioral and cognitive treatments of depression. *Cognitive Therapy and Research, 7,* 111–124.
*Woodward, R., & Jones, R. B. (1980). Cognitive restructuring treatment: A controlled trial with anxious patients. *Behaviour Research and Therapy, 18,* 401–407.
*Wright, J. C. (1976). A comparison of systematic desensitization and social skill acquisition in the modification of social fear. *Behavior Therapy, 7,* 205–210.
*Zane, G., & Williams, S. L. (1993). Performance-related anxiety in agoraphobia: Treatment procedures and cognitive mechanisms of change. *Behavior Therapy, 24,* 625–643.

Received August 1, 1996
Revision received February 10, 1997
Accepted February 27, 1997 ■