Annu. Rev. Psychol. 2001. 52:685–716

EMPIRICALLY SUPPORTED PSYCHOLOGICAL INTERVENTIONS: Controversies and Evidence

Dianne L. Chambless
Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3270; e-mail: chambles@email.unc.edu

Thomas H. Ollendick
Child Study Center, Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0355; e-mail: tho@vt.edu

Key Words empirically supported therapies, empirically validated treatments, psychotherapy efficacy, psychotherapy effectiveness

■ **Abstract** Efforts to increase the practice of evidence-based psychotherapy in the United States have led to the formation of task forces to define, identify, and disseminate information about empirically supported psychological interventions. The work of several such task forces and other groups reviewing empirically supported treatments (ESTs) in the United States, United Kingdom, and elsewhere is summarized here, along with the lists of treatments that have been identified as ESTs. Also reviewed is the controversy surrounding EST identification and dissemination, including concerns about research methodology, external validity, and utility of EST research, as well as the reliability and transparency of the EST review process.

CONTENTS

INTRODUCTION

In 1995, the Task Force on Promotion and Dissemination of Psychological Procedures (henceforth, referred to as Task Force) of Division 12 (Clinical Psychology) of the American Psychological Association (APA) issued the first of three reports [Task Force 1995; Chambless et al 1996, 1998; also available on the Division 12 web page (http://www.apa.org/divisions/div12/est/est.html)] in which it identified a number of psychological interventions as empirically validated treatments [later called empirically supported treatments (ESTs)]. Reaping both praise and opprobrium, these reports have been the subject of conference presentations, newsletter articles, and special sections of journals (e.g. Kendall & Chambless 1998, Elliott 1998b, Kazdin 1996, *Can. Psychol.* 1999). In this review we consider the history of this effort, the evidence amassed by the Task Force and like bodies, and issues that give rise to the controversy over ESTs.

Appointed in 1993, the Task Force was charged with considering issues in the dissemination of psychological treatments of known efficacy. After acceptance by the Division 12 Board of Directors in October of that year, the Task Force's report was circulated to a number of groups for discussion, including the APA Boards of Educational Affairs, Scientific Affairs, and Professional Affairs; the APA Continuing Education Committee, Committee for Approval of Continuing Education Sponsors, and Committee on Accreditation; the Council of University Directors of Clinical Psychology; and the Association of Psychology Postdoctoral and Internship Centers. Individuals were informed of the report and its contents through a symposium at the APA convention, summaries published in *The Clinical Psychologist* and the *APA Monitor* with full copies available upon request (>1000 requests have been filled), and electronic mailings to list servers in clinical psychology. Copies of the full report or of pertinent sections were mailed to the APA Board of Directors, state psychological associations, directors of APA-approved clinical doctoral and internship programs, and those in attendance at an APA conference on postdoctoral education. The report was then published early in 1995, along with responses to comments.

Division 12's interest in promoting the awareness and use of ESTs is part of a broader movement that arose in the United Kingdom and was initially known as evidence-based medicine (Sackett et al 1997). The premises of this movement are that (*a*) patient care can be enhanced by acquisition and use of up-to-date empirical knowledge and (*b*) it is difficult for clinicians to keep up with newly emerging information pertinent to their practice but (*c*), if they do not, their knowledge and clinical performance will deteriorate over the years after their training; consequently, (*d*) clinicians need summaries of evidence provided by expert reviews and instructions on how to access this information during their routine practice. Although developed independently, the Task Force's recommendations are clearly consistent with these tenets. Among the more controversial of its recommendations are that APA engage in an ongoing effort to develop and maintain a list of ESTs for distribution and that training programs in clinical psychology

include some didactic and applied training in ESTs. The latter recommendation followed a survey by task force members (Crits-Christoph et al 1995) of directors of APA-approved clinical-training programs. Respondents (83% of all directors) were asked to indicate which ESTs of a preliminary list assembled by the Task Force were included in students' didactic coursework or practicum training. The authors defined minimal coverage as some inclusion of 25% of ESTs in didactic courses and some clinical training in at least two ESTs. By these criteria, about one-fifth of the programs failed to include minimal coverage of ESTs. More recently, the successor committee to the Task Force has established a web page to disseminate information about ESTs to the public through Division 12's web site (http://www.apa.org/divisions/div12/rev_est/index.shtml).

Efforts to identify psychological ESTs have not been limited to the initial Task Force. Because the first task force was focused largely on ESTs for adults, Division 12 appointed a second task force with an emphasis on ESTs and prevention programs for children, the Task Force on Effective Psychosocial Interventions: A Lifespan Perspective. Reports from this task force have also been published (Spirito 1999, *J. Clin. Child Psychol*. 1998). Division 12 also commissioned a book entitled *A Guide to Treatments That Work*, spearheaded by yet another task force. Edited by Nathan & Gorman (1998), this mammoth book contains reviews of the psychotherapy and pharmacology outcome literature by experts in various psychological disorders. Apart from Division 12, a number of other authors have undertaken reviews of ESTs [e.g. Kendall & Chambless (1998), including publications identifying ESTs in adult, child, marital, and family therapy, Wilson & Gil's review of ESTs for chronic pain (1996), and Gatz et al's review of ESTs for the elderly (1998)]. Psychologists in the United Kingdom have been leaders in this area, including publication of *What Works for Whom?* by Roth & Fonagy (1996), who conducted a review of psychotherapy efficacy for the British National Health Service (Parry 1996). The Canadian Psychological Association's Clinical Psychology Section appointed its own EST task force (see Hunsley et al 1999). Strauss & Kaechele (1998) have described somewhat different but related efforts in Germany.

In the 1995 report, the Task Force published criteria for selection of ESTs and a very preliminary list of 25 treatments that met these criteria. This list was quickly constructed to allow the survey of clinical directors described above and to demonstrate that treatments meeting these criteria could be identified. In subsequent reports, the Task Force expanded the list of ESTs (Chambless et al 1996, 1998) and also collected and published information concerning training opportunities and materials for therapists (Sanderson & Woody 1995, Woody & Sanderson 1998). As of 1998, the list included 71 treatments. The decision by the Committee on Accreditation (American Psychological Association 1996) to include some training in ESTs as part of the guidelines for accreditation of doctoral- and internship-training programs in applied psychology no doubt heightened the already intense interest in the delineation of ESTs. As it became clear that APA would not itself pick up the work of the EST list, Division 12 committed itself to

continuing these efforts by transforming the Task Force into a standing committee charged with evaluating the efficacy and effectiveness of psychological interventions. By a vote of the membership, this transition was approved and took effect in January 1999.

IDENTIFICATION OF EMPIRICALLY SUPPORTED THERAPIES

In Table 1, we summarize the criteria used by the various work groups to define ESTs and, in Tables 2 and 3, their findings to date. The groups are identified by letters: A for the original task force (Chambless et al 1998), B for the special section of *Journal of Pediatric Psychology* (Spirito 1999), C for the special section of *Journal of Child Clinical Psychology* (1998), D for the special section of *Journal of Consulting and Clinical Psychology* (Kendall & Chambless 1998), E for the British review *What Works for Whom?* (Roth & Fonagy 1996), F for Nathan & Gorman's (1998) *A Guide to Treatments that Work*, G for Gatz et al's (1998) review of treatments for the elderly, and H for Wilson & Gil's (1996) review of treatments for chronic pain. For purposes of comparison, we created three rough categories indicating level of support, based on the criteria listed in Table 1. These take into account the number of studies available and their experimental rigor. The following points need to be kept in mind for interpretation of these tables. First, because the different work groups did not use the exact same definitions, the distinction between categories I and II for work groups not following the task force criteria (Chambless et al 1998) completely is not precise. Moreover, Roth & Fonagy (1996) did not distinguish between category-I and -II treatments; thus, their ESTs are listed in the tables under both these categories with a question mark. Second, not all work groups listed promising treatments; those doing so were groups B, E, and F. Third, Nathan & Gorman (1998) eschewed categories. Accordingly, we imposed the category scheme on the reviews in their book, requiring two type-1 studies for inclusion in category I and one type-1 study for inclusion in category II, and relegating treatments supported by only type-2 or -3 studies to category III.

Treatments in category I are supported by at least two rigorous randomized controlled trials (RCTs) showing their superiority to placebo control conditions or another bona fide treatment. Alternatively, work groups following task force criteria (Chambless et al 1998) accepted a large series of rigorous single case experiments as meeting this definition, and authors in the Nathan & Gorman volume (1998) may have accepted the use of a waiting list control group for comparison. Treatments in category II were typically supported by at least one RCT in which the treatment proved superior to a control condition or alternative bona fide treatment. Also, most groups followed the task force criteria in relegating treatments demonstrated to be efficacious only by comparison to waiting list control groups to this category, regardless of the number of supporting studies, and accepted a small series of rigorous single case experiments as meeting threshold for this level

TABLE 1 Workgroup criteria for identification of empirically supported therapies

Division 12 Task Force criteria (Chambless et al 1998), group A
- Well-established treatments
  - I. At least two good between-group design experiments must demonstrate efficacy in one or more of the following ways:
    - A. Superiority to pill or psychotherapy placebo, or to other treatment
    - B. Equivalence to already established treatment with adequate sample sizes
  - OR
  - II. A large series of single-case design experiments must demonstrate efficacy with
    - A. Use of good experimental design and
    - B. Comparison of intervention to another treatment
  - III. Experiments must be conducted with treatment manuals or equivalent clear description of treatment
  - IV. Characteristics of samples must be specified
  - V. Effects must be demonstrated by at least two different investigators or teams
- Probably efficacious treatments
  - I. Two experiments must show that the treatment is superior to waiting-list control group
  - OR
  - II. One or more experiments must meet well-established criteria IA or IB, III, and IV above but V is not met
  - OR
  - III. A small series of single-case design experiments must meet well-established-treatment criteria
- Experimental treatments
  - Treatment not yet tested in trials meeting task force criteria for methodology

Special section of *Journal of Pediatric Psychology* (Spirito 1999) criteria, group B
- Well-established treatments
  - Same as Chambless et al (1998)
- Probably efficacious treatments
  - Same as Chambless et al (1998)
- Promising interventions
  - I. There must be positive support from one well-controlled study and at least one other less-well-controlled study
  - OR
  - II. There must be positive support from a small number of single-case design experiments
  - OR
  - III. There must be positive support from two or more well controlled studies by the same investigator

Special section of *Journal of Clinical Child Psychology* (1998, vol. 27, no. 2) criteria, group C
- Well-established treatments
  - Same as Chambless et al (1998)
- Probably efficacious treatments
  - Same as Chambless et al (1998) except:
    - There must be at least two, rather than one, group design studies meeting criteria for well-established treatments if conducted by the same investigator

Special section of *Journal of Consulting and Clinical Psychology* (Kendall & Chambless 1998) criteria, group D

(*Continued*)

TABLE 1 (*Continued*)

Efficacious and specific
Same as Chambless et al (1998) for well-established treatments
Possibly efficacious and specific treatments
Same as efficacious and specific above except:
Treatment only required to be found superior to rival treatment in one study
Efficacious and possibly specific treatments
Same as efficacious and specific criteria above except:
Treatment was found superior to wait-list group in one study and superior to rival treatment in another study by a different team
Efficacious treatments
Same as Chambless et al (1998) for well-established treatments except:
Treatment must be demonstrated to be better than no treatment but not been shown to be better than nonspecific intervention, placebo, or rival intervention
Possibly Efficacious Treatments
Same as Chambless et al (1998) for probably efficacious treatments

What Works for Whom? (Roth & Fonagy 1996) criteria, group E
Clearly effective treatments
I. There must be a replicated demonstration of superiority to a control condition or another treatment condition
OR
II. There must be a single high-quality randomized control trial in which:
A. Therapists followed a clearly described therapeutic method useable as the basis for training
B. There is a clearly described patient group
Promising limited-support treatments
Treatment must be innovative and a promising line of intervention
OR
Treatment is a widely practiced method with only limited support for effectiveness

A Guide to Treatments That Work (Nathan & Gorman 1998) criteria, group F
Type 1 studies
I. Study must include a randomized prospective clinical trial
II. Study must include comparison groups with random assignment, blind assessments, clear inclusion and exclusion criteria, state-of-the-art diagnostic methods, and adequate sample size for power
III. There must be clearly described statistical methods
Type 2 studies
Clinical trials must be performed, but some traits of type-1 study were missing (e.g. trial with no double blind or group assignment not randomized)
Type 3 studies
I. These are open treatment studies that are aimed at obtaining pilot data
OR
II. These are case control studies in which treatment information was obtained retrospectively

Treatments for older adults (Gatz et al 1998) criteria, group G
Same as Chambless et al (1998) criteria

Treatments for chronic pain (Wilson & Gil 1996) criteria, group H
Same as Chambless et al (1998) criteria

of support. Category III is quite heterogeneous, in that groups E and F accepted a low level of evidence (e.g. a case series) as sufficient evidence to label a treatment as promising, whereas group B demanded more rigorously collected data, such as two or more well-controlled studies conducted by the same investigator.

When ESTs involved the addition of psychological procedures to pharmacological or other medical interventions (e.g. treatment for pain), the control group had to receive the same medical treatments. Work groups varied in their requirements for assessment. For example, those following the task force's criteria (Chambless et al 1998) required only that some reliable and valid method for determination of outcome was employed. Others imposed specific criteria pertinent to a given EST, such as requiring chemical verification of smoking cessation or behavioral observation measures of a child's hyperactive behavior. Given the difficulty in collecting follow-up data unconfounded by additional treatment and extensive attrition, efficacy was typically determined immediately posttreatment.

Table 2 lists ESTs identified for adults, and Table 3 lists those for children. Review of these tables permits some determination of the degrees to which different groups have identified the same treatments in their reviews. When review groups did not use precisely the same label to describe an EST, some judgment was required on our part to determine whether the same treatment was meant. In this case we were guided by the citations authors used to support their categorization of the treatment, as well as our own review of the articles in question.

A number of caveats pertain to the contents of the tables. First, on the whole, review groups have not attempted to determine which treatments are not efficacious. Second, no group essayed a complete review of the psychotherapy literature. For example, the Task Force reports explicitly note that their lists are works in progress. Thus, that a treatment does not appear on any one or all of the lists could have several meanings: (*a*) the treatment in question has fared poorly in research trials, (*b*) the treatment has not been examined in research trials, or (*c*) the treatment was not reviewed. Finally, all groups did not review all treatments. Thus, that one group listed a treatment as an EST and another did not cannot be taken to mean that the groups were in disagreement about a treatment's status. Disagreements may be indicated when one group assigned a treatment to a different category than another. However, because the groups published their work at different times, even then the discrepancy might be due to the publication of new studies.

Despite the number of individuals involved in the various independent review groups and the different approaches taken to evaluation, a review of these tables makes it apparent that, when two or more groups reported a review of the same treatment, the results were generally consistent. This should provide some reassurance to those who have expressed concern about unreliability or biases of individual reviewers for the various work groups. At the time of this review 108 category-I and -II ESTs had been identified for adults and 37 for children, reflecting considerable progress in discernment of ESTs since the first Task Force report with its list of 25 treatments. Nonetheless, that one can identify and disseminate information about ESTs does not address the arguments that it is ill advised to do so. We turn now to a consideration of critiques of the EST work.

TABLE 2 Empirically supported treatments for adults: a summary across workgroups

	Category of Empirical Support[b, c]		
Condition and Treatment[a]	**I**	**II**	**III**
Anxiety and stress			
Agoraphobia/panic disorder with agoraphobia			
CBT	A, E?, F	E?	
Couples communication training as adjunct to exposure		A, D	
Exposure	A, D, E?, F	E?	
Partner-assisted CBT		D, F	
Blood injury phobia			
Applied tension		F	E
Exposure			E
Generalized anxiety disorder			
Applied relaxation	F	A, D, E	
CBT	A, D, E?, F	E?	
Geriatric anxiety			
CBT			F, G
Relaxation		F	
Obsessive-compulsive disorder			
ERP	A, D, E?, F	E?	
Cognitive therapy		A, D	E
RET + exposure			E
Family-assisted ERP + relaxation		D	
Relapse prevention		A	
Panic disorder			
Applied relaxation	F	A, D, E	
CBT	A, D, E?, F	E?	
Emotion-focused therapy			F
Exposure	E?	D, E?	
Post-traumatic stress disorder			
EMDR		A (civilian only), D	
Exposure	F	A, D	
Stress inoculation	F	A, D	
Stress inoculation in combination with CT + exposure	E?	E?, F	
Structured psychodynamic treatment			E
Public-speaking anxiety			
Systematic desensitization		A	
Social anxiety/phobia			
CBT	E?, F	A, D, E?	
Exposure	E?,	A, D, E?, F	
Systematic desensitization		A	

TABLE 2 (*Continued*)

Condition and Treatment[a]	Category of Empirical Support[b, c] I	II	III
Specific phobia			
Exposure	A, E?, F	E?	
Systematic desensitization		A	
Stress			
Stress inoculation	A		
Chemical abuse and dependence			
Alcohol abuse and dependence			
Community reinforcement	E?, F?	A, D, E?, F?	
Cue exposure therapy		A, D	
Cue exposure therapy + urge-coping skills		D	
Cue exposure with inpatient treatment		A	
Motivational interviewing	E?	E?	
BMT + disulfiram	E?, F?	A, D, E?, F?	
Social skills training with inpatient treatment	E?, F?	A, D, E?, F?	
Benzodiazepine withdrawal for panic disorder			
CBT		A	
Cocaine abuse			
Behavior therapy		A	
CBT relapse prevention		A, D	
Opiate dependence			
Behavior therapy (reinforcement)		D	
Brief dynamic therapy		A, D	
CT		A, D	
Depression			
Bipolar Disorder			
Psychoeducation		F	
CBT for medication adherence		F	
Family Therapy			F
Geriatric depression			
Behavior therapy	E?, F	E?, G	
Brief psychodynamic therapy	E?, F	E?, G	
CBT	E?, F	A, E?, G	
Interpersonal therapy		F	
Problem-solving therapy		F, G	
Psychoeducation	F		
Reminiscence therapy (mild-moderate)	F	A, G	
Major depression			
Behavior therapy	A, F	D	
BMT (for those with marital discord)	F	D	
Brief dynamic therapy		A	E

(*Continued*)

TABLE 2 (*Continued*)

Condition and Treatment[a]	Category of Empirical Support[b, c] I	II	III
CBT	A, D, E?, F	E?	
Interpersonal therapy	A, E?, F	D, E?	
Self-control therapy		A, F	
Social problem solving		A, D	
Health problems			
Anorexia			
Behavior therapy	E?	E?	
BFST		F	
CT	E?	E?	
Family therapy			F
Binge-eating disorder			
Behavioral weight control		F	
CBT	F	A	
Interpersonal therapy		A, F	
Bulimia			
CBT	A, E?, F	D, E?	
Interpersonal therapy	E?	A, D, E?, F	
Cancer pain			
CBT			H
Chemotherapy side effects (for cancer patients)			
Progressive muscle relaxation with or without guided imagery		D	
Chronic pain (heterogeneous)			
CBT with physical therapy		A, D, H	
EMG biofeedback		A	
Operant behavior therapy		A, D	
Chronic pain (back)			
CBT	H	A, D	
Operant behavior therapy		D	
Headache			
Behavior therapy	A		
Idiopathic pain			
CBT			H
Irritable-bowel syndrome			
CT		A, D	
Hypnotherapy		D	
Multicomponent CBT		A, D	
Migraine			
EMG biofeedback + relaxation		D	
Thermal biofeedback + relaxation training		A, D	

TABLE 2 (*Continued*)

Condition and Treatment[a]	Category of Empirical Support[b, c] I	II	III
Obesity			
Hypnosis with CBT		A	
Raynaud's			
Thermal biofeedback		A	
Rheumatic disease pain			
Multicomponent CBT	A, D, H		
Sickle cell disease pain			
Multicomponent CT		A	
Smoking cessation			
Group CBT		D	
Multicomponent CBT with relapse prevention	A, D		
Scheduled reduced smoking with multicomponent behavior therapy		A, D	
Somatoform pain disorders			
CBT		F	
Marital discord			
BMT	A, D		
CBT		D	
CT		D	
Emotion-focused couples therapy		A (no more than moderately distressed), D	
Insight-oriented marital therapy		A, D	
Systemic therapy		D	
Sexual dysfunction			
Erectile dysfunction			
Behavior therapy aimed at reducing sexual anxiety and improving communication	E?	E?	
CBT aimed at reducing sexual anxiety and improving communication	E?	E?	
Female hypoactive sexual desire			
Hurlbert's combined therapy		A, D	
Zimmer's combined sex and marital therapy		A, D	
Female orgasmic disorder/dysfunction			
BMT with Masters & Johnson's therapy		D	
Masters & Johnson's sex therapy		A, D	
Sexual-skills training		D	

(*Continued*)

TABLE 2 (*Continued*)

Condition and Treatment[a]	Category of Empirical Support[b, c] I	II	III
Premature ejaculation			
Behavior therapy			E
Vaginismus			
Exposure-based behavior therapy	E?	E?	
Other			
Avoidant personality disorder			
Exposure		F	
Social skills training	E?	E?, F	
Body dysmorphic disorder			
CBT		F	
Borderline personality disorder			
Dialectical behavior therapy	E?	A, E?, F	
Psychodynamic therapy			F
Dementia			
Behavioral interventions applied at environmental level for behavior problems	G		
Memory and cognitive retraining for slowing cognitive decline		G	
Reality orientation		G	E
Geriatric care givers' distress			
Psychoeducation		G	
Psychosocial interventions	E?	E?	
Hypochondriasis			
CBT			F
Paraphilias/Sex Offenders			
Behavior therapy		A	
CBT			F
Schizophrenia			
Assertive case management			F
Behavior therapy and social learning/ token economy programs	F		
Clinical case management			F
CT (for delusions)			E, F
Behavioral family therapy	D, E?, F	A, E?	
Family systems therapy		D	
Social-learning programs	F		
Social-skills training	F	A, D	
Supportive group therapy		F	
Supportive long-term family therapy	D		

TABLE 2 (*Continued*)

Condition and Treatment[a]	Category of Empirical Support[b, c] I	II	III
Training in community living program	F		
Severely mentally ill			
Supported employment		A, F	
Sleep disorders			
Behavior therapy			F
CBT (for geriatric sleep disorders)		G	
Unwanted habits			
Habit reversal and control techniques		A	

[a]CBT, cognitive behavior therapy; BMT, behavioral marital therapy; ERP, exposure plus ritual prevention; BFST, behavioral family systems therapy; EMDR, eye movement desensitization and reprocessing; CT, cognitive therapy; EMG, electromyographic.

[b]Category I, well-established/efficacious and specific/two type-1 studies; Category II, probably efficacious/efficacious/or possibly efficacious/one type-1 study; Category III, promising/type-2 or -3 studies. Only Groups B, E, and F listed Category III treatments.

[c]*Work groups:* A, Task Force (Chambless et al 1998); B, Special section of *Journal of Pediatric Psychology* (Spirito 1999); C, Special section of *Journal of Clinical Child Psychology* (1998); D, Special section of *Journal of Consulting and Clinical Psychology* (Kendall & Chambless 1998); E, *What Works for Whom?* (Roth & Fonagy 1996); F, *A Guide to Treatments That Work* (Nathan & Gorman 1998); G, Gatz et al (1998); H, Wilson & Gil (1996). ?, unclear from author's description whether treatment belongs in Category I or II.

CONTROVERSY

In this section we review a number of the criticisms leveled at the work on ESTs, some of which are particular to the publications of the Task Force and others of which are general to the very concept of identification of ESTs. Much contention stems from guild or economic concerns that the EST findings (*a*) will be misused by managed care companies to disenfranchise practitioners of psychotherapies that are not so designated (e.g. Silverman 1996), (*b*) will make these same practitioners more vulnerable to malpractice suits (Kovacs 1996), or (*c*) will restrict practice to a limited number of treatments, thus precluding flexibility and clinical innovation (see Elliott 1998a). We do not focus on these concerns, which although important, are outside the scientific arena. Rather, in the space permitted, we consider the empirical evidence for and against assertions that the EST list, criteria, and premises are flawed.

Treatment Selection

The Task Force and all other groups listed in Table 1 have adopted a similar approach to their reviews. They have defined ESTs as being specific treatments for specific issues, for example, cognitive-behavior therapy (CBT) for chronic back

TABLE 3 Empirically validated treatments for children and adolescents: a summary across workgroups[a]

	Category of Empirical Support		
Condition and Treatment[b]	**I**	**II**	**III**
ADHD			
Behavioral parent training	C		
Behavior modification in classroom	C		F
Long-term multimodal therapy			E
Anxiety disorders (separation anxiety, avoidant disorder, overanxious disorder)			
CBT		A, C	E
CBT + family AMT		A, C	
Psychodynamic psychotherapy			E
Chronic pain (musculoskeletal disorders)			
CBT			B
Conduct disorder (oppositional defiant disorder)			
Anger control training with stress inoculation (adolescents)		C	
Anger-coping therapy (children)		C	
Assertiveness training		C	
CBT	E?	E?	
Cognitive problem-solving skills	F		
Delinquency prevention program		C	
Functional family therapy	F		
Multisystemic therapy	F	C	
Parent-child interaction therapy		C	
Parent training based on living with children (children)	A, E?, F	C, E?	
Parent training based on living with children (adolescents)	C		
Problem-solving skills training		C	
Rational emotive therapy		C	
Time out plus signal seat treatment		C	
Videotape-modeling parent training	C		
Depression			
Coping with depression course with skills training (adolescents)		C	
CBT (children)		C	
Disruptive disorders			
Structural family therapies			E
Distress due to medical procedures (mainly for cancer)			
CBT	B		
Encopresis			
Behavior modification	E?	A, E?	

TABLE 3 (*Continued*)

Condition and Treatment[b]	Category of Empirical Support I	II	III
Enuresis			
Behavior modification	A, E?	E	
Obesity			
Behavior therapy		A	
Obsessive-compulsive disorder			
ERP			E
Phobias			
CBT		C	
Filmed modeling		C	
Imaginal desensitization		C	
In vivo desensitization		C	
Live modeling		C	
Participant modeling	C		
Rapid exposure (school phobia)	E?	E?	
Reinforced practice	C	A	
Psychophysiological disorder			
Family therapy	E?	E?	
Psychodynamic psychotherapy			E
Pervasive developmental disorders, undesirable behavior in			
Contingency management	E?	E?	
Recurrent abdominal pain			
CBT		D, F	
Recurrent headache			
Biofeedback with self-hypnosis			B
Relaxation/self-hypnosis	B		
Thermal biofeedback		B	

[a]See footnotes in Table 2 for explanation of categories and listing of workgroups.

[b]ADHD, attention deficit/hyperactivity disorder; AMT, anxiety management training.

pain, multisystemic family therapy for conduct disorder, or brief psychodynamic therapy for depression. Copious evidence already exists for the proposition that psychotherapy in general is effective for clients in general [see, e.g., the meta-analyses by Shapiro & Shapiro (1982) and Weisz et al (1995)]. Thus, the various task forces and independent groups have sought to take the next step, to identify more precisely which treatments practitioners of evidence-based psychological interventions might consider for particular clients. This decision in and of itself has proved controversial (e.g. Garfield 1996).

Sample Description

Arguing that it is unwise to conclude that all treatments are beneficial for all types of clients, the Task Force required that tests of ESTs be conducted on a sample identified with some specificity. This might be and often was a diagnosis based on some system such as the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association 1994) but could include any reliable and valid approach to defining the sample of interest, for example, a cutoff score on a questionnaire indicating martial distress for couples seeking marital therapy. A number of critics have objected to this decision for two reasons: (*a*) they dispute that there are meaningful differences in the efficacy of different forms of psychological interventions (e.g. Wampold et al 1997), and (*b*) they decry the acceptance of the medical model implied in frequent use of *Diagnostic and Statistical Manual of Mental Disorders* diagnoses to describe samples (e.g. Henry 1998). We consider the first of these objections in a later section (see "Treatment Specificity" below). Regarding the second objection, critics have argued that no system of describing clients is needed, that groupings of clients are impossible because each case is unique, or that diagnosis is dehumanizing (e.g. Bohart et al 1998). Others have called for a different sort of organizing principle, perhaps one based on important personality factors that may affect response to psychotherapy, such as reactance (e.g. Beutler et al 1996), or on functional relationships (i.e. the cause of the dysfunctional behavior; e.g. Scotti et al 1996). In response, task force members (Task Force 1995) have noted that some method for describing clients is necessary to enable clinicians to evaluate the likelihood of generalization from research samples for their own practice and to organize data. Without some categorization, synthesis of evidence is extremely difficult, if not impossible. Fonagy & Target (1996) noted that, whatever the limitations of diagnostic systems such as the DSM, critics have yet to suggest a better feasible alternative.

Research Design

A second criticism of the criteria concerns the Task Force's reliance on RCTs (or comparable single-case design approaches) to test the outcome of research on psychological interventions. This criterion has been attacked from three perspectives. First, it has been rejected as a poor approach to gaining knowledge by those who suggest that qualitative research would be more appropriate (e.g. Bohart et al 1998). Second, because cognitive and behavioral approaches have been more often the subject of RCTs than have other forms of psychotherapy, some fear that the criteria lend unfair advantage to CBT over other therapies (e.g. Bohart et al 1998). Third, Seligman (1995), among others, has argued that, because clients in treatment in the community are not randomly assigned to treatment, ESTs based on RCTs are unlikely to generalize to clinical practice. Only the last of these points can be addressed through empirical findings, and we devote a separate section to these later (see "Effectiveness" below).

The Task Force's decision to require careful definition of the independent variable (the therapy or therapies), most often, although not necessarily, in the form

of a treatment manual, has resulted in heated controversy. A treatment manual is a statement of the principles and procedures of a psychological intervention, including both prescribed and proscribed interventions. Manuals may be in the detailed form of session-by-session guidelines or may provide extensive flexibility within the general framework provided by the authors (Kendall et al 1998). The better manuals are richly elaborated with examples of dialogue illustrating application of the procedures and with descriptions of courses of action to take when problems arise in treatment. The Task Force concluded that treatment manuals or their equivalent in the form of a clear description of the treatment are necessary to provide an operational definition of the intervention under study, because very different procedures may fall under the general rubric of, for example, psychodynamic psychotherapy or CBT. Failing such a description, clinicians are unable to determine the similarity of their own practices to ESTs, and educators are ignorant of how they might train their students in ESTs. That is, in terms of generalizable knowledge, it is meaningless to say that a treatment works without being able to say what that treatment is.

In part, the negative reaction to the requirement that EST research include manuals or other clear descriptions of the treatment appears to be an emotional one (see Fonagy 1999), with some clinicians reacting to a perceived threat to their independence of practice. For example, some have viewed manuals as "promoting a cookbook mentality" (Smith 1995, p. 40), "paint[ing] by numbers" (Silverman 1996, p. 207), or even as "a hangman of life" (Lambert 1998, p. 391). This reaction is captured nicely by Bohart et al (1998), who wrote, "[the therapist is]. . .a disciplined improvisational artist, not a manual-driven technician" (p. 145). Others have viewed manualization (as it has come to be called) in more positive terms. For example, Wilson (1998, p. 363) has asserted that ". . . use of standardized, manual-based treatments in clinical practice represents a new and evolving development with far-reaching implications for the field of psychotherapy." Indeed, in a survey of licensed psychologists, Addis & Krasnow (2000) found that practitioners were equally likely to be positively vs negatively inclined toward treatment manuals.

Seligman (1995) and Garfield (1996), among others, have objected to the use of manuals on another basis—that therapists in the community do not follow manuals and, thus, their use in efficacy research limits generalization. Indeed, according to Addis & Krasnow's (2000) survey, 47% of licensed psychologists in practice had never used manuals in their clinical work. However, that therapists may not, on the whole, use manualized treatments now does not address the question of whether it would be desirable for them to do so. At present, 23% of psychologists in practice report that they have never heard of treatment manuals, and of those aware of their existence, 38% are unclear as to what manuals are (Addis & Krasnow 2000). For a further review of the controversy surrounding manualization, see Addis et al (1999).

Some authors (e.g. Elliott 1998a) have complained that the task force criteria are too lenient, because they are focused solely on the efficacy of the treatments in RCTs and not on their effectiveness, that is, on whether these treatments work in what the authors term real clinical practice. This issue, in our judgment, is far more

important than whether therapists are currently using ESTs as described in treatment manuals in their practice. We consider this topic at some length below (see "Effectiveness"). Finally, there are general complaints about the clarity of the criteria and the methods for reaching decisions (e.g. Elliott 1998a); for example, when are two versions of CBT for depression similar enough to lump together and when should they be considered separate treatments? The Task Force required that experiments have been conducted with "good" methodology, but what defines good? Chambless & Hollon (1998) have provided considerable detail on evaluation of the methodology of psychotherapy research as used by the Task Force and others (Kendall & Chambless 1998). Despite this detail, these authors did not draw hard and fast decision rules, leaving room for potential bias on the reviewers' part. In our view, this is one of the more telling criticisms of the Task Force's work. The group provided no evidence of the reliability of its decisions and no detailed explanation of the process involved in reaching those decisions. Nonetheless, as previously noted, a review of Table 2 indicates that multiple review groups have generally arrived at the same conclusions about the efficacy of treatments evaluated. (Treatments in Table 3 are less likely to have been reviewed by more than one group.)

Cognizant of the importance that the reviews should be viewed as reasoned and impartial, the Division 12 Committee on Science and Practice, which succeeded the Task Force, has postponed the EST review process for several years in favor of developing an elaborate manual for members to use in evaluating treatments. The twelfth draft of this manual was recently completed (JR Weisz & KM Hawley, unpublished manuscript) and is under review by committee members. The present plan is for this manual to be used by two other groups identifying ESTs, the Committee for Empirically Supported Practice of APA Division 53 (Clinical Child Psychology) and the Interdisciplinary Committee on Evidence-Based Youth Mental Health Care. These groups provided feedback on the development of the manual. It is likely that treatments already on the EST list will be reevaluated according to the criteria set forth in this new manual. Because the Committee has yet to vote on final approval of the manual, we do not summarize it in detail here. However, from our own review we conclude that adoption and use of this manual will address most complaints about clarity of process, as well as many of the technical concerns.

Standardization of Treatment

The use of treatment manuals in EST research and in training of psychotherapists has been criticized not only on emotional grounds but also on the grounds that use of treatment manuals will lead to a degradation in the quality of treatment that clients receive (e.g. Garfield 1996, Henry 1998). The assumptions are (*a*) that therapists trained in a standardized treatment will not be able to deviate flexibly from such protocols when necessary to treat a particular case and (*b*) that a standardized treatment approach will be less beneficial than a treatment program designed specifically for an individual case. Concerning the first point, Garfield, Henry,

and others have pointed to research on adherence to protocol in psychotherapy research to buttress their arguments that manual-based training will be detrimental to the efficacy of psychotherapy. The empirical findings on this point are inconsistent, with some authors finding that, at least under some circumstances (e.g. rupture of the therapeutic alliance), greater adherence is related to poorer outcome (e.g. Castonguay et al 1996), whereas others have found greater adherence to be related to better outcome (e.g. Frank et al 1991). However, such arguments miss the point, in that such research tells us nothing about the success of manual-trained therapists in their unstructured clinical practice. We know of no research on this first point, and other authors (e.g. Moras 1993) have argued that students grasp the general principles of treatment and acquire skills more quickly when they are presented in the organized form of a manual rather than solely relying on the informal teaching that occurs during supervision. Thus, research directed at this question would be highly desirable.

For now, this leaves us to question the second point: Is standardized treatment less efficacious than an individualized treatment program? We located only three studies addressing this issue: (*a*) Jacobson et al's (1989) comparison of standardized behavioral marital therapy to the same form of therapy delivered flexibly with regard to the total length of treatment and the amount of time spent on any one component; (*b*) Schulte et al's (1992) comparison of treatment of phobias (mostly agoraphobia) by standardized treatment with prolonged exposure vs individually designed treatment in which the therapist was given total flexibility in designing the treatment for each client; and (*c*) Emmelkamp et al's comparison (1994) of self-directed exposure plus ritual prevention for obsessive-compulsive disorder to a program of various cognitive and behavioral treatments based on an individual behavioral analysis. In all of these studies, the same therapists administered treatments in the standardized vs individualized conditions, ruling out the possibility that therapist main effects contributed to the outcome, and clients were randomly assigned to treatments.

In none of the studies was flexible treatment statistically superior to standardized treatment. Indeed, Schulte et al (1992) found the standardized treatment for phobia to be significantly better than that based on the therapist's discretion, largely because some therapists failed to use exposure treatment in this condition, even though exposure is the best validated treatment for phobia. On the other hand, in the marital-therapy study, there was a trend at 6-month follow-up in favor of the flexible treatment group, and effect sizes were higher in the flexible group. Unfortunately, because length of treatment was confounded with flexibility of treatment procedures (treatment was, on average, longer in the flexible condition), these data are difficult to interpret. In addition, in this study therapists were constrained to use the same treatment components regardless of condition. Thus, this study is less a test of manualization than of a more vs less flexible manual. Additional research of this nature with a wider variety of problems and including children and adolescents should be conducted. For now, the data do not support the assertion that manualizing treatments will be detrimental to clients.

Treatment Specificity

Obviously, if all forms of psychotherapy were equally efficacious for all problems, the identification of ESTs would be a much less important exercise. Many would claim that reviews of psychotherapy research have demonstrated this to be the case. For example, Wampold et al (1997) conducted a meta-analysis in which they tested the hypothesis that the effect size for comparisons between different types of psychotherapy is zero. Their results were consistent with this hypothesis and are typical of findings from similar meta-analyses. A common interpretation of such findings is that nonspecific factors (e.g. the quality of the therapist-client relationship) are responsible for all of the meaningful variance in treatment outcome. Nonspecific factors are clearly of great importance in psychotherapy, with, for example, the therapeutic relationship accounting for about 9% of the outcome variance [see meta-analysis by Horvath & Symonds (1991)]. Nonetheless, we would argue that it is hasty to conclude that there are no meaningful differences in the effects of psychotherapies, for several reasons.

First, meta-analyses such as those conducted by Wampold et al (1997) are generally focused on a somewhat restricted sample of clients, including college students and outpatients with problems with anxiety, depression, eating disorders, and the like. Comparisons among different theoretical schools are not available for treatment of the chronically mentally ill and the developmentally disabled. We doubt that such comparisons will be conducted because of the strength of the belief, right or wrong, that behavioral methods are the treatments of choice for these clients.

Second, meta-analyses demonstrating lack of specificity often include little research on children or adolescents. For example, Wampold et al (1997) did not search journals in the child area such as the *Journal of Clinical Child Psychology*, the *Journal of Abnormal Child Psychology*, or the *Journal of the American Academy of Child and Adolescent Psychiatry*. Meta-analyses of treatment research on children and adolescents have yielded different conclusions. In their meta-analysis of treatment outcome studies conducted between 1967 and 1993 with children and adolescents, Weisz et al (1995) concluded that behavioral treatments (e.g. behavioral contracting, parent management training, modeling, and CBT) were more effective than nonbehavioral treatments (e.g. insight-oriented therapy or client-centered counseling). Casey & Berman (1985) reached a similar conclusion in their early meta-analysis of psychotherapy research with children, conducted between 1952 and 1982.

Third, it is often suggested that some problems such as depression are benefited by a wide range of treatments (Beutler et al 2000), whereas others, such as obsessive-compulsive disorder and phobias, respond better to specific behavioral treatments or CBTs than to other treatments. To evaluate these arguments, we now examine evidence for the specificity of psychotherapy effects, taken from studies meeting criteria for EST research. This body of literature is very large. Owing to space considerations, we include only a sampling of studies for two problems for

which specificity research is available—anxiety disorders and depression. Within these categories we focus, whenever possible, on particularly well-designed studies in which authors have ruled out alternative explanations for differences between treatments, such as uncontrolled therapists' effects and differences in the quality of the therapeutic relationship and in the credibility of treatment.

Anxiety Disorders Research on treatment of adult obsessive-compulsive disorder has demonstrated a high degree of specificity, in that different forms of behavior therapy have significantly different effects. Fals-Stewart et al (1993), Lindsay et al (1997), Hodgson et al (1972), and Marks et al (1975) all compared exposure plus response prevention, a well-established EST, to progressive muscle relaxation, either alone or elaborated as anxiety management training. Designs in the 1970s were quasi-experimental, but the more recent two studies were RCTs. In all four studies, exposure plus response prevention was significantly superior to the alternative treatment. Lindsay et al (1997) tested whether differences in the therapeutic relationship might account for these findings. This proved not to be the case. When interviewed by an independent assessor, clients rated therapists as highly and comparably supportive and understanding, regardless of treatment condition. Unfortunately, these authors did not assess clients' expectancies for change or credibility of treatment. This oversight was corrected in a study of a different anxiety disorder.

Studying generalized anxiety disorder among adults, Borkovec & Costello (1993) found that CBT was significantly superior to nondirective therapy at posttest and 1-year follow-up. This study represents one of the most sophisticated comparisons available in the specificity literature. The authors demonstrated that clients in CBT and nondirective therapy were comparable in their ratings of expectancy of change and their therapists. Furthermore, in both conditions therapists were guided by detailed treatment manuals, and the authors verified that treatment effects were consistent across therapists. Consistent with the theory of nondirective therapy, observers rated clients in that condition as having greater depth of experiencing than clients in the other conditions. Nonetheless, these clients failed to improve as much as those in CBT.

Research on treatments for children's and adolescent's anxiety disorders also demonstrates specificity. Using a controlled, single-case design, Ollendick (1995) showed that a developmental adaptation of Barlow & Craske's (1994) CBT for panic disorder with agoraphobia was beneficial for adolescents, whereas education and support alone were not. Treatment gains persisted at 6-month follow-up. In the specific phobia domain, specificity is also evident (Ollendick & King 1998, 2000). For example, participant modeling (in which the therapist models, prompts, and assists the child in approaching the phobic stimulus) has been shown to be more effective than nondirective, supportive therapy as well as other variants of behavioral and cognitive therapies, including systematic desensitization and filmed modeling. Furthermore, reinforced practice (based on principles of shaping, positive reinforcement, and extinction) has been shown to be more effective than verbal coping skills and therapist modeling alone.

Depression Specificity is also evident in psychotherapy effects for depression among children and adolescents. For example, Stark et al (1991) compared CBT to nondirective, supportive therapy in the treatment of children who reported high levels of depressive symptoms on the Children's Depression Inventory. At postintervention and at 7-month follow-up, children in the CBT group reported fewer depressive symptoms on a semistructured interview and indicated fewer depressive symptoms on the Children's Depression Inventory than children in the supportive condition. Brent et al (1997) obtained consistent results, showing that CBT was more effective than supportive, nondirective therapy with adolescents who have major depressive disorders. Moreover, CBT was marginally more effective than a family intervention utilizing techniques of functional family therapy.

Specificity breaks down where treatment of depressed adults is concerned. For example, in the large-scale National Institute of Mental Health (NIMH) study comparing CBT to interpersonal therapy (Elkin et al 1989), no significant differences between treatments were observed. Similarly, Thompson et al (1987) found no evidence of treatment specificity in their comparison of behavioral, cognitive, and brief dynamic therapies for depressed elders, and Jacobson et al (1996) found that the full CBT package was no more efficacious than one of its components, behavioral activation (i.e. assignment of activities to build mastery and positive mood), alone.

It is not entirely clear why the patterns of findings differ between depressed children and adults. This may be the result of developmental differences. It is also possible that the difference between the findings for adults and children has to do with the therapies with which CBT is contrasted. In the child studies, the comparison group received supportive, nondirective therapy, whereas, in the adult studies, comparison treatments were behavior therapy and psychodynamic therapy.

Effectiveness

One of the strongest arguments posed against the dissemination of ESTs is the dearth of evidence that research treatments are beneficial when applied (*a*) in ordinary clinical settings such as private practices and community mental health and medical centers and (*b*) to ordinary clients, that is, those who are not selected according to an extensive list of exclusion criteria and those who have not agreed to be randomly assigned to treatments as part of a research enterprise. In the psychotherapy literature, this distinction is known as the difference between evidence of efficacy vs effectiveness [see Hoagwood et al (1995) or Moras (1998) for a concise presentation of these concepts]; almost always this distinction overlaps highly with the contrast between experiments with high internal vs high external validity. It is true that, until recently, psychotherapy researchers concentrated almost exclusively on efficacy research, but effectiveness data are beginning to emerge. Here we describe samples of effectiveness research on a number of well-established treatments, including, for adults, CBT for depression and panic disorder and, for children, two parent training programs for conduct/oppositional defiant disorder.

Panic Disorder In a study notable for its external validity, Wade et al (1998) reported the effectiveness of Barlow & Craske's (1994) CBT for panic disorder with or without agoraphobia for 110 clients of a community mental health center. Therapists were staff members trained to administer a standard 15-session protocol. Minimal exclusion criteria were employed. Clients showed significant pretest-posttest changes on all major outcome measures, and improvements were comparable in effect size to changes observed with the same treatment program in efficacy studies. However, the authors reported that the dropout rate was higher than that in efficacy studies, although still considerably lower than the dropout rate usually observed in community mental health center samples.

A significant challenge to the external validity of efficacy studies is the limited range of ethnic groups included in research samples. It is typical for the large majority of subjects to be white or for the ethnic characteristics of samples to go unreported. Sanderson et al (1998) reported the results of 12 sessions of Barlow & Craske's (1994) CBT for panic disorder when applied to a medical center outpatient clinic sample of 30 low income, urban clients, the majority of whom were Latino. Statistically significant improvement was observed, although clients remained more impaired than those in the benchmark study. The authors noted that the level of disorder in their population was initially more severe but also suggested that the high rate of stress in this sample may have limited the success of treatment.

Depression We located five studies of the effectiveness of CBT for depression, all with positive results. Peterson & Halstead (1998) examined the outcome of an abbreviated form of standardized CBT (6 sessions vs the typical 20) administered in a group format. Clients were 210 patients referred to a depression management program in military clinics. Minimal exclusion criteria were applied. Significant pretest-posttest reductions on the Beck Depression Inventory were observed, regardless of the initial level of severity of depression. The authors noted that the average reduction in depressive symptoms (37%) was less than that usually reported in efficacy studies (57%), but they also pointed out that their treatment was considerably briefer for practical reasons—cost reduction and prevention of attrition.

Persons and colleagues have published the results of two investigations of CBT for depression in private-practice samples (Persons et al 1988, 1999). Treatment was applied in a flexible manner according to individual case formulations to clients with a variety of depressive disorders regardless of their comorbidity. Significant improvement was observed, and outcomes for intention-to-treat samples (including all clients, regardless of whether they completed treatment) were comparable to those in benchmark efficacy studies. However, private-practice clients received, on average, more sessions than those in efficacy studies. Whether improvement would have been comparable had this not been the case couldn't be determined from these data.

In two separate samples, Organista et al (1994) and Miranda et al (J Miranda, F Azocar, K Organista, E Dwyer, P Arean, manuscript under review) tested the

benefits of standardized group CBT for depression in low-income medical outpatients treated at a hospital depression clinic. The majority were members of ethnic minority groups (Latino or African-American) and had significant medical problems. Significant improvement was observed in both studies, and Miranda et al reported that 62%–80% (depending on ethnicity) no longer met diagnostic criteria for depression after treatment. Nonetheless, on average clients remained in the depressed range, according to questionnaire scores. The dropout rate was higher with this disadvantaged population (40%–60%) than in efficacy studies [e.g. 32% (Elkin et al 1989)], but change was significant for the intention-to-treat sample (i.e. the sample of all clients who began treatment, including dropouts). Miranda et al found that adding case management services reduced dropouts significantly.

Oppositional-Defiant Disorder Two effectiveness studies have been conducted with children presenting with oppositional problems. In the first study, Taylor et al (1998) examined the effectiveness of Webster-Stratton's Parents and Children Series (PACS) at a community-based children's mental health center. Parents seeking help in managing their 3- to 8-year-old-children's behaviors were randomly assigned to one of three conditions: PACS, treatment as usual (TAU), or a waiting-list control group. TAU was defined by each therapist and included ecological, solution-focused, family systems, and popular-press parenting approaches. Therapists were affiliated with the center; 7 were trained in Webster-Stratton's approach, and 11 were not. Both active treatments were superior to the waiting list, but mothers in the PACs program reported significantly fewer behavior problems and greater satisfaction with treatment than mothers in the TAU program. Outcomes were similar to those obtained in the efficacy trials. Thus, this trial provides evidence of PACS's specificity as well as effectiveness. However, because therapists were not crossed with treatment conditions or randomly assigned to conditions, it is also possible that PACS's superiority to TAU was due to some superiority of PACS therapists unrelated to the treatment.

A second, uncontrolled effectiveness trial with oppositional children was recently reported. Using parent management training and child social skills training, Tynan et al (1999) treated 55 consecutive admissions to a hospital-affiliated child psychiatry clinic. The children, ranging in age from 5 to 11 years, all presented with oppositional problems, and most were comorbid with other disorders including attention deficit/hyperactivity disorder. Therapists were trained in the implementation of the manualized treatment. Pre-post changes were large, significant, and comparable to benchmark efficacy studies.

Focus of Treatment

A further criticism leveled at the external validity of EST research is the charge that, in such studies, the researchers focus only on so-called symptoms. Depending on theoretical orientation, clinicians may reject the importance of symptoms in favor of goals such as self-transcendence (Kovacs 1996) or improvement in overall life adjustment. To call problems such as depression or panic "symptoms" is

consistent with approaches such as psychodynamic theory, in which it is assumed that these conditions reflect some more central underlying problems. In behavioral or cognitive approaches, these conditions are generally not viewed as symptoms but as valid problems in their own right. Be that as it may, it is legitimate to question whether ESTs lead to improvement in quality of life or overall adjustment.

In response to calls for data on quality of life, EST researchers are now including measures for this construct in their assessment battery, but these studies are mostly in progress, with findings not yet available (Gladis et al 1999). In the adult arena, we located two studies, including three different ESTs. No such studies were found in the child and adolescent arena. Indeed, Gladis et al pointed out that adequate instruments for assessing quality of life in children have yet to be developed. For the adult studies, we examined findings at both posttest and follow-up, because changes in overall adjustment may be slow to emerge, typically lagging after changes in presenting problems even in non-symptom-focused treatment, such as psychodynamic psychotherapy (e.g. Howard et al 1993).

Cognitive Therapy for Depression Blatt et al (2000) analyzed data from the NIMH Treatment for Depression Collaborative Research Program, in which two ESTs—cognitive and interpersonal therapies—were compared to treatments with an antidepressant plus clinical management and with a placebo plus clinical management. Analyses controlled for overall symptom severity and level of functioning at pretest, as well as other variables that had proved predictive of treatment outcome. There were no differences among treatments at posttest, but by an 18-month follow-up, clients in both psychotherapy groups rated their life adjustment significantly more positively than clients in the placebo or medication group. Findings were not due to differences in expectancy.

Cognitive Behavioral Therapy for Panic Disorder Telch et al (1995) compared the effects of 12 sessions of CBT to a delayed-treatment control group. Quality of life measures included assessment of global impairment in work, social and leisure activities, and family life. At posttest, the treated clients showed significantly more improvement than waiting-list clients on both measures of global adjustment, as well as on most of the measures' subscales. Because the control group participants entered treatment after posttest, follow-up data are uncontrolled. Six months after treatment, client scores continued to be significantly improved compared with pretest scores, with no significant change from posttest to follow-up.

CONCLUSION

In conclusion, we summarize the arguments for and against identification and dissemination of ESTs and our responses:

1. Argument: ESTs should be ignored because this effort has been the work of a small group of biased individuals in Division 12 (Society for Clinical Psychology of the American Psychological Association).

Response: Tables 2 and 3 summarize the results of searches by eight different review groups. Because the groups had different foci (e.g. geriatric vs child clients), their review areas often did not overlap. When they did, there is evidence of considerable agreement in the judgment of different groups. This suggests that, once decisions are based on common assumptions (e.g. use of RCTs, specifically defined samples, and treatments), EST designations can and have been made reliably. This assertion needs to be tempered by the presence of some overlap in the membership of the Task Force and the authors of the *Journal of Consulting and Clinical Psychology* special section on this subject (Groups A & D) (Kendall & Chambless 1998).

2. Argument: Quantitative research is not an appropriate paradigm for psychotherapy research. Qualitative research or clinical observation should be the evidence source.

   Response: Evidence reviewed in this chapter cannot address this question, which results from a schism in paradigms. No matter how large or consistent the body of evidence found for identified ESTs, findings will be dismissed as irrelevant by those with fundamentally different views, and such views characterize a number of practitioners and theorists in the psychotherapy area.

3. Argument: EST research is based on treatment manuals or their equivalent, and use of manuals to train therapists will lead to decrements in the quality of psychotherapy.

   Response: We found no studies addressing the concern that, when working outside a restricted treatment protocol, manual-trained therapists would be less effective than therapists trained without manuals. Such research would be valuable in addressing these concerns. We did locate two studies in which the results from therapists trained in ESTs were compared under two conditions: one in which therapists were free to design individually tailored treatments (all within a general cognitive-behavioral framework) and one in which therapists operated under the standard EST guidelines. Standardized treatment proved equivalent or superior to individually tailored treatment. Unfortunately, both of these studies involved treatments of anxiety disorders. Additional research in this area with a broader range of treatments and problems is clearly needed.

4. Argument: There is no difference in efficacy among various forms of psychotherapy; hence, identification of ESTs is unnecessary.

   Response: In our abbreviated review of anxiety disorders and childhood depression, we found considerable evidence of specificity, even within cognitive and behavioral methods. Authors were not consistent in the care they took to rule out alternative explanations for differences among treatments, such as possible confounding discrepancies in expectancy and the therapeutic relationship. However, in studies in which such confounds could be discounted, treatment differences were still obtained.

Nonetheless, evidence of specificity is not uniform, in that, for adult depression, no treatment has been clearly demonstrated to be superior to another. These data and those from meta-analyses indicate that the question of specificity is a complex one and may depend on the target problem and clients' age (child vs adult).

5. Argument: EST research should be ignored because it will not generalize to clinical practice, defined as typical settings, clients, therapists, and treatment goals.

   Response: Thus far, the evidence reviewed indicates that ESTs are effective in clinical settings and with a diverse group of clients. Nonetheless, this conclusion needs to be tempered by the following considerations. (*a*) As would be expected in effectiveness research, these studies were high in external validity but low on internal validity, typically relying on uncontrolled pretest-posttest designs and evaluation by comparison with standard outcomes in the efficacy literature. (*b*) Clients in effectiveness studies often did not improve as much as those in efficacy studies or did not achieve as positive a level of functioning at the end of treatment. (*c*) Clients in some effectiveness studies received more treatment than is typical in RCTs, in the form of a longer course of treatment or adjunctive treatments such as concomitant pharmacotherapy or case management services. (*d*) The body of effectiveness research is small.

Hence, although we found no evidence that ESTs could not be beneficially applied in real clinical settings by those trained or supervised in their use, depending on the population served, treatments may need to be modified or included as part of a broader course of treatment. More research on effectiveness is clearly a priority if clinicians are to give credence to the value of ESTs. The NIMH is currently making such studies a priority in its mission for treatment research; this should ensure that effectiveness studies rapidly multiply.

Where treatment goals are concerned, we found very little research on whether ESTs affect quality of life. The two studies we found yielded positive results. To the degree that ESTs focus on clients with significant psychological disturbance, researchers are likely to find that improvements in presenting problems will be associated with enhanced quality of life. This follows from the substantial correlations of measures of disorder with quality of life measures. These effect sizes are typically moderate to large (Gladis et al 1999, Telch et al 1995). Should EST researchers focus on healthy but dissatisfied clients, we would expect quality-of-life measures to play a more unique role and, indeed, to become the central outcome measures in such research. More research on improved quality of life as an effect of ESTs is clearly needed, and such data should be available soon from studies presently under way.

The drive to identify and disseminate ESTs is less than a decade old. From the review above, it is clear that there is much to be learned about the application of ESTs and their benefits in clinical practice. Whatever the reluctance of some

to embrace ESTs, we expect that the economic and societal pressures on practitioners for accountability will encourage continued attention to these treatments (cf Beutler 1998). At the same time, it is important to note that the effective practice of evidence-based psychotherapy involves more than the mastery of specific procedures outlined in EST manuals. Almost all ESTs rely on therapists' having good nonspecific therapy skills. For example, in the NIMH Treatment of Depression Collaborative Research Program of two ESTs (interpersonal therapy and cognitive therapy) for depression, in which therapists followed elaborate treatment manuals, treatment effects varied according to therapists' competence (Shaw et al 1999). Client characteristics such as ability to form an alliance with the therapist (Krupnick et al 1996) and initial functioning (e.g. Elkin et al 1995) also proved to be important in predicting treatment outcome. Research on the interaction of client characteristics (e.g. personality) with treatment approaches is in its early stages (e.g. Beutler et al 1991). Such research is difficult to conduct because of the large sample sizes required for sufficient power. Yet such research has great potential for addressing practitioners' questions about which among several ESTs might be best for a particular client. Thus, the practice of evidence-based psychotherapy is a complex one, and ESTs are only one piece of the puzzle.

ACKNOWLEDGMENTS

We are very grateful to Steffany Fredman and Carol Woods for their invaluable assistance in the construction of Tables 1 through 3.

LITERATURE CITED

Addis ME, Krasnow AD. 2000. A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. *J. Consult. Clin. Psychol.* 68:331–39

Addis ME, Wade WA, Hatgis C. 1999. Barriers to dissemination of evidence-based practices: addressing practitioners' concerns about manual-based psychotherapies. *Clin. Psychol. Sci. Pract.* 4:430–41

Am. Psychiatr. Assoc. 1994. *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: Am. Psychiatr. Assoc. 886 pp. 4th ed.

Am. Psychol. Assoc. 1996. *Guidelines and Principles for Accreditation of Programs in Professional Psychology*. Washington, DC: Am. Psychol. Assoc.

Barlow DH, Craske MG. 1994. *Mastery of Your Anxiety and Panic*, Vol. II. Albany, NY: Graywind. 150 pp.

Beutler LE. 1998. Identifying empirically supported treatments: What if we didn't? *J. Consult. Clin. Psychol.* 66:113–20

Beutler LE, Clarkin J, Bongar B. 2000. *Guidelines for the Systematic Treatment of the Depressed Patient*. Oxford, UK: Oxford Univ. Press. 455 pp.

Beutler LE, Engle D, Mohr D, Daldrup RJ, Bergan J, et al. 1991. Predictors of differential response to cognitive, experiential, and self-directed psychotherapeutic procedures. *J. Consult. Clin. Psychol.* 59:333–40

Beutler LE, Kim JE, Davison E, Karno M,

Fisher D. 1996. Research contributions to improving managed health care outcomes. *Psychotherapy* 33:197–206

Blatt SJ, Zuroff DC, Bondi CM, Sanisolow CA III. 2000. Short and long-term effects of medication and psychotherapy in the brief treatment of depression: further analyses of data from the NIMH TDCRP. *Psychother. Res.* 10:215–34

Bohart AD, O'Hara M, Leitner LM. 1998. Empirically violated treatments: disenfranchisement of humanistic and other psychotherapies. *Psychother. Res.* 8:141–57

Borkovec TD, Costello E. 1993. Efficacy of applied relaxation and cognitive behavioral therapy in the treatment of generalized anxiety disorder. *J. Consult. Clin. Psychol.* 61:611–19

Brent DA, Holder D, Kolko D, Birmaher B, Baugher M, et al. 1997. A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. *Arch. Gen. Psychiatry* 54:877–85

Can. Psychol. 1999. Symposium: empirically supported treatments in psychology. 40:289–327

Casey RJ, Berman JS. 1985. The outcome of psychotherapy with children. *Psychol. Bull.* 98:388–400

Castonguay LG, Goldfried MR, Wiser S, Raue PJ, Hayes AM. 1996. Predicting the effect of cognitive therapy for depression: a study of unique and common factors. *J. Consult. Clin. Psychol.* 64:497–504

Chambless DL, Baker M, Baucom DH, Beutler LE, Calhoun KS, et al. 1998. Update on empirically validated therapies, II. *Clin. Psychol.* 51(1):3–16

Chambless DL, Hollon SD. 1998. Defining empirically supported therapies. *J. Consult. Clin. Psychol.* 66:7–18

Chambless DL, Sanderson WC, Shoham V, Bennett Johnson S, Pope KS, et al. 1996. An update on empirically validated therapies. *Clin. Psychol.* 49(2):5–18

Crits-Christoph P, Frank E, Chambless DL, Brody C, Karp JF. 1995. Training in empirically-validated treatments: What are clinical psychology students learning? *Prof. Psychol. Res. Pract.* 26:514–22

Elkin I, Gibbons RD, Shea MT, Sotsky SM, Watkins JT, Pilkonis PA. 1995. Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J. Consult. Clin. Psychol.* 63:841–47

Elkin I, Shea T, Watkins JT, Imber SD, Sotsky SM, et al. 1989. National Institute of Mental Health Treatment of Depression Collaborative Research Program: general effectiveness of treatments. *Arch. Gen. Psychiatry* 46:971–82

Elliott RE. 1998a. Editor's introduction: a guide to the empirically supported treatments controversy. *Psychother. Res.* 8:115–25

Elliott RE, ed. 1998b. The empirically supported treatments controversy. *Psychother. Res.* 8:115–70 (special issue)

Emmelkamp PMG, Bouman T, Blaauw E. 1994. Individualized versus standardized therapy: a comparative evaluation with obsessive-compulsive patients. *Clin. Psychol. Psychother.* 1:95–100

Fals-Stewart W, Marks AP, Schafer J. 1993. A comparison of behavioral group therapy and individual behavior therapy in treating obsessive-compulsive disorder. *J. Nerv. Ment. Dis.* 181:189–93

Fonagy P. 1999. Achieving evidence-based psychotherapy practice: a psychodynamic perspective on the general acceptance of treatment manuals. *Clin. Psychol. Sci. Pract.* 6:442–44

Fonagy P, Target M. 1996. Should we allow psychotherapy research to determine clinical practice? *Clin. Psychol. Sci. Pract.* 3: 245–50

Frank E, Kupfer DJ, Wagner EF, McEachran AB, Cornes C. 1991. Efficacy of interpersonal psychotherapy as a maintenance treatment of recurrent depression: contributing factors. *Arch. Gen. Psychiatry* 48:1053–59

Garfield SL. 1996. Some problems with "validated" forms of psychotherapy. *Clin. Psychol. Sci. Pract.* 3:218–29

Gatz M, Fiske A, Fox LS, Kaskie B, Kasl-Godley JE, et al. 1998. Empirically validated psychological treatments for older adults. *J. Ment. Health Aging* 41:9–46

Gladis MM, Gosch EA, Dishuk NM, Crits-Christoph P. 1999. Quality of life: expanding the scope of clinical significance. *J. Consult. Clin. Psychol.* 67:320–31

Henry WP. 1998. Science, politics, and the politics of science: the use and misuse of empirically validated treatments. *Psychother. Res.* 8:126–40

Hoagwood K, Hibbs E, Brent D, Jensen P. 1995. Introduction to the special section: efficacy and effectiveness in studies of child and adolescent psychotherapy. *J. Consult. Clin. Psychol.* 63:683–87

Hodgson R, Rachman S, Marks IM. 1972. The treatment of chronic obsessive-compulsive neurosis: follow-up and further findings. *Behav. Res. Ther.* 10:181–89

Horvarth AO, Symonds DB. 1991. Relationship between working alliance and outcome in psychotherapy: a meta-analysis. *J. Couns. Psychol.* 38:139–49

Howard KI, Lueger RJ, Maling MS, Martinovich Z. 1993. A phase model of psychotherapy outcome: causal mediation of change. *J. Consult. Clin. Psychol.* 61:678–85

Hunsley J, Dobson KS, Johnston C, Mikail SF. 1999. Empirically supported treatments in psychology: implications for Canadian professional psychology. *Can. Psychol.* 40:289–302

Jacobson NS, Dobson KS, Truax PA, Addis ME, Koerner K, et al. 1996. A component analysis of cognitive-behavioral treatment for depression. *J. Consult. Clin. Psychol.* 64:295–304

Jacobson NS, Schmaling KB, Holtzworth-Munroe A, Katt JL, Wood LF, et al. 1989. Research-structured vs. clinically flexible versions of social learning-based marital therapy. *Behav. Res. Ther.* 27:173–80

J. Clin. Child Psychol. 1998. Special issue: empirically supported psychosocial interventions for children. 27:138–226

Kazdin AE, ed. 1996. Validated treatments: multiple perspectives and issues. *Clin. Psychol. Sci. Pract.* 3:216–67 (special issue)

Kendall PC, Chambless DL, eds. 1998. Empirically supported psychological therapies. *J. Consult. Clin. Psychol.* 66:3–167 (special issue)

Kendall PC, Chu B, Gifford A, Hayes C, Nauta M. 1998. Breathing life into a manual: flexibility and creativity with manual-based treatments. *Cogn. Behav. Pract.* 5:177–98

Kovacs AL. 1996. Winter: "We have met the enemy and he is us!" *AAP Adv.*, pp. 6, 19, 20, 22

Krupnick JL, Sotsky SM, Simmens S, Moyer I, Elkin I, et al. 1996. The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: findings in the NIMH Treatment of Depression Collaborative Research Program. *J. Consult. Clin. Psychol.* 64:532–39

Lambert MJ. 1998. Manual-based treatment and clinical practice: hangman of life or promising development? *Clin. Psychol. Sci. Pract.* 5:391–95

Lindsay M, Crino R, Andrews G. 1997. Controlled trial of exposure and response prevention in obsessive-compulsive disorder. *Br. J. Psychiatry* 171:135–39

Marks IM, Hodgson R, Rachman S. 1975. Treatment of chronic obsessive-compulsive neurosis by in-vivo exposure: a two-year follow-up and issues in treatment. *Br. J. Psychiatry* 127:349–64

Moras K. 1993. The use of treatment manuals to train psychotherapists: observations and recommendations. *Psychotherapy* 30:581–86

Moras K. 1998. Internal and external validity of intervention studies. In *Comprehensive Clinical Psychology*, ed. AS Bellack, M Hersen, 3:201–24. Oxford, UK: Elsevier. 380 pp.

Nathan PE, Gorman JM, eds. 1998. *A Guide to Treatments That Work*. New York: Oxford Univ. Press. 594 pp.

Ollendick TH. 1995. Cognitive-behavioral treatment of panic disorder with agoraphobia in adolescents: a multiple baseline design analysis. *Behav. Ther.* 26:517–31

Ollendick TH, King NJ. 1998. Empirically supported treatments for children with phobic and anxiety disorders: current status. *J. Clin. Child Psychol.* 27:156–67

Ollendick TH, King NJ. 2000. Empirically supported treatments for children and adolescents. In *Child and Adolescent Therapy*, ed. PC Kendall, pp. 386–425. New York: Guilford. 436 pp. 2nd ed.

Organista KC, Munoz RF, Gonzalez G. 1994. Cognitive-behavioral therapy for depression in low-income and minority medial outpatients: description of a program and exploratory analyses. *Cogn. Ther. Res.* 18:241–59

Parry G. 1996. *NHS Psychotherapy Services in England, Rep. 96PP0043*. London: Dep. Health. 112 pp.

Persons JB, Bostrom A, Bertagnolli A. 1999. Results of randomized controlled trials of cognitive therapy for depression generalize to private practice. *Cogn. Ther. Res.* 23:535–48

Persons JB, Burns DB, Perloff JM. 1988. Predictors of dropout and outcome in cognitive therapy for depression in a private practice setting. *Cogn. Ther. Res.* 12:557–75

Peterson AL, Halstead TS. 1998. Group cognitive behavior therapy for depression in a community setting: a clinical replication series. *Behav. Ther.* 29:3–18

Roth AD, Fonagy P. 1996. *What Works for Whom? A Critical Review of Psychotherapy Research*. New York: Guilford. 484 pp.

Sackett DL, Richardson WS, Rosenberg W, Haynes RB. 1997. *Evidence-Based Medicine*. New York: Churchill Livingstone. 250 pp.

Sanderson WC, Raue PJ, Wetzler S. 1998. The generalizability of cognitive behavior therapy for panic disorder. *J. Cogn. Psychother.* 12:323–30

Sanderson WC, Woody SR. 1995. Manuals for empirically validated treatments: a project of the Task Force on Psychological Interventions. *Clin. Psychol.* 48(4):7–11

Schulte D, Kunzel R, Pepping G, Schulte-Bahrenberg T. 1992. Tailor-made versus standardized therapy of phobic patients. *Adv. Behav. Res. Ther.* 14:67–92

Scotti JR, Morris TL, McNeil CB, Hawkins RP. 1996. DSM-IV and disorders of childhood and adolescents: Can structural criteria be functional? *J. Consult. Clin. Psychol.* 64:1177–91

Seligman MEP. 1995. The effectiveness of psychotherapy: the Consumer Reports study. *Am. Psychol.* 50:965–74

Shapiro DA, Shapiro D. 1982. Meta-analysis of comparative therapy outcome studies: a replication and refinement. *Psychol. Bull.* 92:581–604

Shaw BF, Elkin I, Yamaguchi J, Olmsted M, Vallis TM, et al. 1999. Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *J. Consult. Clin. Psychol.* 67:837–46

Silverman WH. 1996. Cookbooks, manuals, and paint-by-numbers: psychotherapy in the 90's. *Psychotherapy* 33:207–15

Smith EWL. 1995. A passionate, rational response to the "manualization" of psychotherapy. *Psychother. Bull.* 30:36–40

Spirito A, ed. 1999. Empirically supported treatments in pediatric psychology. *J. Pediatr. Psychol.* 24:87–174 (special issue)

Stark K, Rouse L, Livingston R. 1991. Treatment of depression during childhood and adolescence: cognitive-behavioral procedures for the individual and family. In *Child and Adolescent Therapy*, ed. PC Kendall, pp. 165–206. New York: Guilford. 436 pp.

Strauss BM, Kaechele H. 1998. The writing on the wall: comments on the current discussion about empirically validated treatments in Germany. *Psychother. Res.* 8:158–77

Task Force Promot. Dissem. Psychol. Procedures. 1995. Training in and dissemination of empirically-validated psychological

treatments: report and recommendations. *Clin. Psychol.* 48(1):3–23

Taylor TK, Schmidt F, Pepler D, Hodgins C. 1998. A comparison of eclectic treatment with Webster-Stratton's parents and children series in a children's mental health setting: a randomized controlled trial. *Behav. Ther.* 29:221–40

Telch MJ, Schmidt NB, Jaimez L, Jacquin KM, Harrington PJ. 1995. Impact of cognitive-behavioral treatment on quality of life in panic disorder patients. *J. Consult. Clin. Psychol.* 63:823–30

Thompson LW, Gallagher E, Steinmetz Breckenridge J. 1987. Comparative effectiveness of psychotherapies for depressed elders. *J. Consult. Clin. Psychol.* 55:385–90

Tynan WD, Schuman W, Lampert N. 1999. Concurrent parent and child therapy groups for externalizing disorders: from the laboratory to the world of managed care. *Cogn. Behav. Pract.* 6:3–9

Wade WA, Treat TA, Stuart GL. 1998. Transporting an empirically supported treatment for panic disorder to a service clinic setting: a benchmarking strategy. *J. Consult. Clin. Psychol.* 66:231–39

Wampold BE, Mondin GW, Moody M, Stich F, Benson K, et al. 1997. A meta-analysis of outcome studies comparing bona fide psychotherapies: empirically, "All must have prizes." *Psychol. Bull.* 122:203–15

Weisz JR, Weiss B, Han SS, Granger DA, Morton T. 1995. Effects of psychotherapy with children and adolescents revisited: a meta-analysis of treatment outcome studies. *Psychol. Bull.* 117:450–68

Wilson GT. 1998. Manual-based treatment and clinical practice. *Clin. Psychol. Sci. Pract.* 5:363–75

Wilson JJ, Gil KM. 1996. The efficacy of psychological and pharmacological interventions for the treatment of chronic disease-related and non-disease-related pain. *Clin. Psychol. Rev.* 16:573–97

Woody SR, Sanderson WC. 1998. Manuals for empirically supported treatments: 1998 update. *Clin. Psychol.* 51(1):17–21

Annual Review of Psychology
Volume 52, 2001

CONTENTS