Course: SW 6302
Instructor: Byrnes

Title: Meta-Analysis of Comparative Therapy Outcome Studies
Author: Shapiro, David

Psychological Bulletin
1982, Vol. 92, No. 3, 581-604

SW-6302-B

0033-2909/82/9203-0581$00.75

Byrnes F2

Meta-Analysis of Comparative Therapy Outcome Studies: A Replication and Refinement

David A. Shapiro and Diana Shapiro
Medical Research Council/Social Science Research Council Social and Applied Psychology Unit
University of Sheffield, Sheffield, England

The results are reported of a meta-analysis of 143 outcome studies, published over a 5-year period, in which two or more treatments were compared with a control group. Consistent with previous reviews, the mean of the 1,828 effect size measures obtained from the 414 treated groups approached one standard deviation unit, and differences among treatment methods accounted for, at most, 10% of the variance in effect size. The impact of differences between treatment methods was outweighed by the combined effects of other variables, such as the nature of the target problem under treatment, aspects of the measurement methods used to assess outcome, and features of the experimental design. However, multiple regression analysis suggested that differences between treatments were largely independent of these other factors. Direct comparisons between pairs of treatments figuring together in the same subsets of the data suggested some consistent differences, with cognitive and certain multimodal behavioral methods yielding favorable results. The practical implications of the conclusions drawn were limited, however, by the predominantly analogue nature of the research reviewed and its unrepresentativeness of clinical practice.

Recent reviews of the comparative psychotherapy outcome literature (Frank, 1979; Luborsky, Singer, & Luborsky, 1975; Smith & Glass, 1977; Smith, Glass, & Miller, 1980), together with the results of the widely acclaimed study by Sloane, Staples, Cristol, Yorkston, and Whipple (1975), have converged on the conclusion that diverse therapies are modestly, but equally, effective. On the other hand, proponents of behavioral therapies, such as Eysenck (1978), Kazdin and Wilson (1978), and Rachman and Wilson (1980), have denounced this conclusion as being based on misconceived aggregation of data from unsound research studies and contradicted by the results of specific, well-conducted studies alleged to favor certain behavioral methods (Bandura, 1977; Franks & Wilson, 1978; Rachman & Hodgson, 1980).

Meta-analysis, which was used by Smith and Glass (1977) and Smith et al. (1980), is a quantitative technique in which the results for each dependent variable in each study are expressed as a difference between means for treated and control groups divided by the standard deviation of the control group scores. These effect size scores are then averaged across studies, and the impact on effect size of several study characteristics (such as treatment method, client and therapist variables, and measurement and design features) is determined empirically. Beyond the psychotherapy outcome field, this method has been applied to such issues as the effect of school class size on achievement (Glass & Smith, 1979); the relation between social class and achievement (White, 1982); the effects of cooperative, competitive, and individualistic goal structures on achievement (Johnson, Maruyama, Johnson, Nelson, & Skon, 1981); the validity of Fiedler's contingency model of leadership effectiveness (Strube & Garcia, 1981); interpersonal expectancy or experimenter effects (Rosenthal & Rubin, 1978); self-serving bias in inter-

This study was presented at the annual meeting of the Society for Psychotherapy Research, Aspen, Colorado, June 1981.

Diana Shapiro is now with the North Derbyshire District Psychology Service.

We gratefully acknowledge helpful discussions with Chris Barker, Gene Glass, Paul Jackson, Mary Lee Smith, and Peter Warr.

Requests for reprints should be sent to David A. Shapiro, MRC/SSRC Social and Applied Psychology Unit, The University, Sheffield S10 2TN, England.

personal influence situations (Arkin, Cooper, & Kolditz, 1980); sex differences in conformity (Cooper, 1979); and sex differences in influencibility (Eagly & Carli, 1981). Proponents of meta-analysis, such as Glass (1978) and Smith (1980), claim that it is less subjective and more precise than the conventional, qualitative literature review or the tallying of positive, neutral, and negative results in the box-score method used by Luborsky et al. (1975). On the other hand, critics of meta-analysis argue that its superiority may be more apparent than real (Barber, 1978; Elashoff, 1978; Eysenck, 1978; Gallo, 1978; Kazdin & Wilson, 1978; Mansfield & Bussey, 1977; Rachman & Wilson, 1980).

The contribution of meta-analysis to literature reviews received a balanced appraisal from Cook and Leviton (1980). These authors applaud meta-analysis for its utility in reviewing large sets of data, its criticism of poor practices in qualitative reviews, its ability to establish dependable generalizations, and its potential to offer clues to the explanation of findings. Nonetheless, they caution that meta-analysis does not obviate the need for "judgements about the definition of the area of investigation, the relevance of methodological and substantive characteristics of studies, and the appropriate meta-analytic tools to be used" and that it can "lead to an unwarranted psychological sense of security if there is a consistent replication of a relationship." Meta-analysis thus requires the assumption that its estimates are unbiased or the demonstration via subsidiary analyses that "no bias of importance resulted from the principal forces from which bias would be expected" (Cook & Leviton, 1980).

The present study was designed to replicate the Smith and Glass (1977) comparison among treatment methods, using meta-analytic methods refined in the light of criticisms advanced by Kazdin and Wilson (1978) and Rachman and Wilson (1980). These criticisms are discussed by Shapiro and Shapiro (1982). Our basic question is this: Do the results of recent controlled comparisons between treatment methods support the assertion of equivalent effects, as propounded by Smith and Glass (1977), or is there evidence for the contrary view that certain behavioral methods are superior (Kazdin & Wilson, 1978; Rachman & Wilson, 1980)? Amon the refinements embodied in the presen study were the following:

1. Rachman and Wilson (1980) com plained that many of the studies included b Smith and Glass (1977) contained no contr group. Furthermore, interpretation of severa of the simultaneous comparisons betwee behavioral and nonbehavioral methods o which these authors base their equivalenc argument depends on the use of control-ref erencing procedures, whereby hypothetica control-group scores are derived from th results of comparable studies in the meta analysis (Smith et al., 1980). Strictly speak ing, the absence of an untreated contr group does not invalidate an otherwise well designed comparison between two activ treatments. Since our concern was with bot absolute and relative efficacy of differen treatment methods, it was deemed cost ef fective to follow the preferences of Rachma and Wilson (1980) and consider only studie including untreated or minimally treate controls.

2. A related criticism of the Smith an Glass (1977) study was the excessive relianc on the aggregation of data from disparat studies. Although Smith and Glass (1977 achieved statistical control over discernibl sources of bias via regression analysis an experimental control via the eliminatio from a subanalysis of all studies not involvin direct comparisons, we preferred to restric our attention to studies making simultaneou comparisons between two or more treat ments and a control group.

3. Smith and Glass (1977) omitted severa behavioral studies considered important b Rachman and Wilson (1980). Rather tha attempt an exhaustive analysis, we took a a representative sample all published, con trolled comparisons between treatments lo cated via *Psychological Abstracts*, 1975–1979 Although no search procedure is infallible we are confident that any omissions did no seriously distort our findings.

4. Rachman and Wilson (1980) criticized the inclusion of unpublished doctoral disser tations in the Smith and Glass (1977) anal ysis. In addition to the possibility that such studies may not meet the criteria of accept ability for publication, their relative inacces

sibility creates problems for the critic wishing to consult the source data. We therefore excluded dissertations from the analysis.

5. Smith and Glass (1977) were criticized by Rachman and Wilson (1980) for their failure to differentiate between different kinds of outcome measurement and for their novel conclusion that behavioral methods tended to be assessed by more subjective outcome measures. We incorporated refinements in the categories and dimensions used to characterize outcome measurement in an attempt to overcome these objections.

It should be noted, however, that some of the criticisms of meta-analysis cannot be directly met by refining the methods. In particular, the claim that the source studies are simply too disparate to permit generalization, of which much is made by critics such as Eysenck (1978), Kazdin and Wilson (1978), and Rachman and Wilson (1980), is properly addressed not to the meta-analytic method as such but rather to anyone seeking to generalize from the data in question (Shapiro & Shapiro, 1982).

Our goal in the present study was to contribute to the debate about generalizing from existing literature on comparative therapy outcomes by ascertaining the extent to which conclusions may be drawn that meet the requirements of Cook and Leviton (1980). Whatever the scientist's reservations about such generalization, it is nonetheless relied on when preferred orientations for therapy training and practice are selected on the basis of outcome literature (Shapiro, 1980; Shapiro & Shapiro, 1977).

Method

Literature Search

Definition of area of study. We followed Smith and Glass (1977) in using Meltzoff and Kornreich's (1970) definition of psychotherapy:

> The informed and planful application of techniques derived from established psychological principles, by persons qualified through training and experience to understand these principles and to apply these techniques with the intention of assisting individuals to modify such personal characteristics as feelings, values, attitudes and behaviors which are judged by the therapist to be maladaptive or maladjustive. (p. 6)

We interpreted this definition to exclude studies in which no clear target problem or complaint could be identified. Studies of clients below the age of 16 years were excluded. Analogue research was included. No rational criterion exists to distinguish absolutely analogue from clinical research (Kazdin, 1978; Mathews, 1978; Smith & Glass, 1977), so characteristics contributing to this alleged distinction, such as problem severity and duration of therapy, were entered as variables in the analysis. Data on process variables bearing no direct relation to client status with respect to the target problem or general well-being were excluded.

Design requirements. All studies followed the conventional group comparative design, including at least three groups. In every study, two or more groups received different psychological treatments, and another group (a no-treatment or, failing that, a minimal-treatment group) could be appropriately designated the control group for the calculation of effect sizes. Designs dependent on own-control or within-subjects comparisons were excluded, except where data from the first treatment block permitted between-subjects comparisons. Measures lacking data from the control group (such as follow-up data not obtained from these subjects or obtained after these subjects had received treatment) were excluded.

Reporting requirements. All included studies were published reports listed in *Psychological Abstracts*, 1975–1979. Dissertations and other documents with restricted circulation were excluded. All studies presented sufficient data to permit the calculation or estimation of effect sizes. The minimum requirement here was the provision of *ns* and *p* values. No reference was made to unpublished supporting documents. Brief reports were included, although these had more missing data and required more estimation of effect sizes than most extended reports.

Search implementation. The large number of therapy titles encountered precluded a dependable computer search (for studies comparing two or more treatments with a control). This was therefore undertaken manually by the second author, using *Psychological Abstracts* indexes. In the initial location of papers, an effort was made to err in the direction of inclusion. Thus, titles and available abstracts led to the identification of some 400 studies; on closer examination, however, only 143 studies, reported in the 140 papers listed in the Appendix, proved suitable for inclusion. In order to avoid bias, no use was made of reviews or bibliographies. Any errors of omission in the search procedure were thus unlikely to bias our results. Of the 143 studies, 21 (15%) were also included by Smith et al. (1980); the two meta-analyses are largely independent of one another. The two meta-analyses also cover different, although overlapping, time periods. Smith et al. (1980) included studies appearing in any year up to 1977, whereas publication dates of the present source studies ranged from 1974 to 1979.

Development and implementation of coding system. The coding system was developed from that used by Smith and Glass (1977), of which details were kindly supplied by Glass, amended and extended to meet criticisms and possible limitations. All coding was carried out by the two authors, both clinical psychologists. The first author has 10 years' postdoctoral research and clinical experience; the second, 5 years' research and 2 years' clinical experience after a master's-level clinical training.

Table 1
Treatment Method and Effect Size

Method	No. of groups	No. of studies	No. of comparisons	Effect size	*SD*	*SE*	Advantag
Behavioral	310	134	56	1.06	.84	.05	.32**
Rehearsal, self-control, and monitoring	38	21	16	1.01	.90	.15	.20
Biofeedback	9	9	9	.91	.48	.16	−.33
Covert behavioral	19	13	10	1.52	1.23	.28	.22
Flooding	18	10	9	1.12	.73	.17	.11
Relaxation	42	31	27	.90	.76	.12	−.14
Systematic desensitization	77	55	50	.97	.65	.07	.04
Reinforcement	28	17	13	.97	.71	.13	.36
Modeling	11	8	6	1.43	1.18	.36	.07
Social skills training	14	14	14	.85	.71	.19	.13
Study skills training	4	4	4	.26	3.9	.20	−.75
Cognitive	35	22	20	1.00	.53	.09	.40***
Dynamic/humanistic	20	16	13	.40	.33	.07	−.53**
Mixed (mainly behavioral)	40	28	24	1.42	1.03	.16	.52**
Unclassified (mainly behavioral)	18	14	14	.78	.59	.13	−.23*
Minimal	41	36	36	.71	.73	.11	−.56***

* $p < .05$. ** $p < .01$. *** $p < .001$.

Most coding operations were refined after pilot coding of some 20 studies, which were recoded with the final system. Three codings, however, were made only after relevant information had been obtained from almost all the studies and categories devised on the basis of this information. These codings were of treatment method, target problem, and measurement technology. In developing the system and in the coding itself we constantly took care to avoid reference to outcome data, so as to prevent bias. All effect-size calculations and estimations were performed either subsequent to the coding of a study on other variables or by the author not responsible for coding the study in other respects.

Independent coding of 20 studies by both authors resulted in at least 80% agreement on each coding dimension.

Treatment and therapist variables. Treatments were classified into 15 categories, as shown in Table 1. *Behavior rehearsal, self-control, and self-monitoring* explicitly entailed active self-direction by the client, in contrast with the allegedly passive conditioning paradigms of desensitization and reinforcement methods (D'Zurilla & Goldfried, 1971; Goldfried, 1971; Kanfer & Karoly, 1972; McFall & Twentyman, 1973). *Biofeedback, flooding, systematic desensitization,* and *modeling* were all defined in the conventional manner of the behavioral literature (Kazdin & Wilson, 1978). *Covert behavioral* methods included covert sensitization (Cautela, 1967), covert reinforcement, and covert modeling but excluded flooding and desensitization, even if wholly imaginal. *Relaxation* included meditation procedures as well as conventional relaxation based on Western rationales. *Reinforcement* methods included both positive and aversive conditioning using overt stimuli. *Social skills training* often included elements similar to treatments coded as behavior rehearsal, self-control, and self-monitoring but also involved a broader range of target-relevant a tivities than could readily be subsumed within the latt category. *Cognitive* methods included the cognitive an cognitive-behavioral methods associated with the nam of Ellis (1962), Beck (1976), and Meichenbaum (1977 packages including these methods alongside conve tional behavioral methods were classified as cognitiv since these methods in any case include some elemen of conventional behavior therapy. *Dynamic and h manistic* methods included therapies based on psych dynamic, client-centered, or other humanistic rati nales, including encounter and gestalt. *Mixed* treatmen were methods that defied classification into any one the categories because they contained elements of mo than one, and *unclassified* methods were too dissimil to the existing categories to belong in any one but t infrequently represented to form a category of their ow Both of the latter groups were predominantly behavior in orientation.[1]

Subsidiary treatment variables. Also coded were t *mode* of therapy (individual, group, couples/families, mixed), *therapist experience* (in years, with undergra uates coded 0, postgraduate trainees coded 2 unless sta of training was specified, and doctoral level therapis coded 5 unless length of subsequent experience cou be inferred), and *duration of therapy* (in hours, with session counted as 1 hour unless less than 40 minute Finally, the *reproducibility* of the treatment method w

[1] Supplementary documentation, comprising tabl listing the source studies in which each treatment occu together with the frequency distributions, means, a standard errors obtained at each level of each code, obtainable from JSAS *Catalogue of Selected Documen in Psychology,* 1982, *12,* 46. (Ms. No. 2506)

coded. The maximum code, 3, required a full description given in the text or available from the author or another cited work, or reference to a manual or direct replication of another author's well-described method. The intermediate code of 2 was assigned for brief descriptions of method or reference to adaptation of a method well described elsewhere. The lowest code, 1, was assigned where the description was minimal and imprecise.

Client variables. The presenting *target problem* was recorded for every study, and these targets were grouped into 5 *target classes* (see Table 2). Where it could be positively made, *psychiatric diagnosis* of the client groups was also recorded. Rather than assume that all minor disorders such as specific phobias and performance anxieties should be classified as neurotic, psychiatric diagnosis was only recorded when offered by the author of the source study. Although Smith and Glass (1977) also coded client intelligence and client–therapist similarity, examination of source studies suggested that the only relevant data available in most cases is the *clients' educational level* (low/mentally retarded, average, high, or mixed). Mean client *age* was also coded. We also coded two important aspects of the analogue–clinical continuum. *Severity/screening* was coded on a 3-point scale (Clinical severity = 3, 80th percentile on target measure or thorough screening = 2, below 80th percentile or inadequate screening = 1), and the *source* of clients was classified as solicited, committed, or sought treatment.

Contextual variables. The *setting* of treatment (laboratory, student health, paraprofessional, clinical, or home) was recorded, together with available information on the presence or absence of *concurrent medication.*

Outcome measurement. Outcome measures were coded in four respects in place of the two codings applied by Smith and Glass (1977). The Smith and Glass (1977) "type of outcome measure" coding was renamed *domain*, in accordance with its reference to the area of functioning (fear/anxiety, self-esteem, adjustment, work/school achievement, personality traits, social behavior, emotional/somatic disorders, and physiological stress) tapped by the measure. For summary analysis, domain was recorded in terms of the assumed tractibility of the functions assessed, with fear/anxiety and self-esteem coded 1, and all others coded 0. The Smith and Glass (1977) dimension of reactivity was differentiated into two concepts. Relatedness to the goals of treatment, considered a virtue by behavioral researchers, was coded separately on a 4-point scale of *specificity* (projective measures = 1, standardized trait measures = 2, directly related measures or physiological measures where target is nonphysiological = 3, behavioral tests or physiological measures of physiological targets = 4). The remaining procedural aspects of fakeability and liability to the influences of demand characteristics were coded on a reduced, 4-point scale of *reactivity*, on which therapist ratings were coded 4, simple self-reports and behavior in the presence of a nonblind experimenter were coded 3, blind ratings and standardized tests were rated 2, and physiological and achievement measures were coded 1. An aspect of measurement insufficiently examined by Smith and Glass (1977) was the type of *measurement technology* employed. This was coded on a 4-point soft–hard scale, whereby self-ratings and projective tests were coded 1, psychometric measures and behavioral ratings were coded 2, behavioral counts were coded 3, and real-life performance measures and physiological data were coded 4.

Design quality and internal validity. Aspects of internal validity included *mode of assignment of clients* (coded on a 4-point scale representing increasing degrees of control: nonrandomized = 1, randomized without matching = 2, randomized with matching of groups = 3, randomized with matching sets = 4), *assignment of therapists* (categorized as for clients but with further categories involving a single therapist in all groups and counterbalancing therapists across all groups), *attrition rates* in treated and control groups (coded for each measure separately), *pretreatment equivalence* on each outcome measure (no significant difference between groups on any measure = 3, no difference on measure in question but other measures indeterminate or showing differences = 2, differences on measure in question = 1), and degree of *blindness* of the person obtaining the outcome data to the treatment received by the client (single blind = 3, knew group composition = 2, acted as therapist = 1). Other design variables included *sample size* of both treatment and control groups, and *number of therapists* involved in the treatment group.

Effect size calculation. The methods of calculation and estimation followed those applied by Smith et al. (1980, Appendix 7), kindly made available prior to publication by Glass. The method used to obtain the effect size was coded. Wherever possible, mean difference over control standard deviation was used. Failing sufficient data for this, error terms were calculated from statistics such as F, t, Neuman-Keuls q, and so on. Failing even this amount of information, a conservative estimate was obtained from supplied ns and p values; these estimates were more accurate and less conservative if means were supplied. Also coded in this category were a few cases in which effect sizes were calculated by assigning numerical values to frequency categories reflecting at least three levels of outcome. Dichotomous frequency data were submitted to simple prohibit analysis (Finney, 1972). Also coded was the *source of data* (posttest scores, covariance adjusted posttest, prepost difference scores, or other). Finally, we also recorded the number of null effect sizes for each treatment as a partial check for spurious inflation of effect size by selective reporting. We recorded the number of measures mentioned by the authors on which data were not available because they failed to differentiate treated from control groups, and for which effect sizes hence were not calculated. Intermediate cases were not included in this tally because although the source paper indicated that treated groups were at least marginally superior to controls, effect sizes could not be estimated.

Data reduction. Consideration of each outcome measure and each occasion of measurement yielded varying numbers of effect size measures for each treatment group. Smith et al. (1980) reduced their data by eliminating redundant measures; that is, measures matching others in measurement characteristics and effect size or subtest scores on multifactorial instruments that could be averaged without apparent loss of information. We preferred to retain all data, except that measures permitting only a relatively imprecise effect size estimate

were discarded if the study provided more complete data on enough other measures, thus yielding a more accurate indication of effect size. Analyses confined to parameters that were invariant across all effect sizes obtained for a given treatment group were conducted, with each treatment group represented only once, by the mean of all effect sizes obtained for that group.

Analysis

Data analysis used Release 8.0 of SPSS (Statistical Package for the Social Sciences; Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975) and proceeded through the phases outlined below. Issues relating to the application of inferential statistics to meta-analysis are discussed by Glass, McGaw, and Smith (1981), Shapiro and Shapiro (1982), and Strube and Hartmann (1982).

Overall effect size. Descriptive statistics were obtained for the complete sample of effect sizes, for comparison with Smith and Glass's (1977; Smith et al., 1980) estimates.

Description of the data base. The data of the study were characterized in terms of distributions on all the coded parameters. This provided a profile of the comparative outcome literature undertaken during the period covered by the study and thus indicated the likely generalizability of its findings.

Correlates of effect size. Univariate breakdowns and correlations were obtained to ascertain the relationship of each parameter with effect size. The parameters were conceptualized as falling into five groups, comprising treatment and therapist variables, client variables, contextual variables, measurement variables, and design variables.

Comparative evaluations of treatment methods. To overcome the confounding influence of other variables covarying across studies with treatment method, the mean effect size obtained for each treatment was compared with the mean effect size of all other treatments found in the studies involving that treatment. This comparison was performed as follows, for each treatment method in turn. First, the data from all studies involving the given treatment were selected. Then advantage scores were obtained from these data by calculating, for each of the selected studies, a mean effect size for the treatment in question, and a mean effect size for all other treatments combined. The latter mean was subtracted from the former for each study in turn, and the mean of these difference scores was evaluated via a *t* test. These calculations were performed twice, the second calculation involving the prior elimination of data from minimal treatments.

Controlled comparisons between treatment methods. A similar method was applied to controlled comparisons between treatment methods. To ensure dependable results, comparisons between two methods were only considered if they were performed in at least four source studies and yielded at least 10 effect sizes for each of the two treatments. After data from all studies making a given comparison were selected, differences were evaluated following similar methods to those used for comparative evaluation of single methods. A mean effect size was obtained for each of the two treatments in each study, and *t* tests were used to evaluate the difference from zero of the resultant study-by-study difference scores.

Treatment × Client interactions. Cross-tabulation o effect size data in terms of the broad classes of treatmen type and target class was evaluated via two-way analysi of variance. Interaction effects from such analysis woul serve as evidence of differential effectiveness of treat ment approaches with different types of presenting prob lems.

Multiple regression analysis. In order to assess th extent to which treatment methods influence outcom independently of other study parameters, sets of vari ables representing treatment, client, measurement, an design variables were entered sequentially in a series o regression equations, with effect size as the dependen variable. These analyses served to supplement the lim ited number of direct comparisons between treatmen methods that were available by comparing the impac on effect size of different treatments while controllin statistically for the influence of other study parameter without restricting the analysis to direct controlled com parisons.

Results

Overall Effect Size

The mean value of the 1,828 effect size was .93, and the standard deviation, 1.16 The mean value implies that the averag treated client lies at the 82nd percentile o untreated clients. If all the 540 null effec sizes recorded are assumed to take the valu zero, then the overall mean effect size is re duced to .72. Omission of the 177 effect size (M = .57) associated with minimal or pla cebo treatment raises the overall mean effec size from .93 to 1.03. Of the 1,828 effect size 206 (11.3%) were negative. The 414 mea effect sizes, of which one represented eac of the 414 treatment groups in the stud yielded a mean of .98 and a standard devia tion of .80.

Description of the Data Base

Associated with each of the 1,828 sepa rately coded effect size measures were score on the parameters outlined above. Parame ters that were by definition constant for a effect size data obtained on a single treatmer group were analyzed on a one-score-pe group basis over the 414 groups included i the study. Statistical analysis of the 1,82 effect size measures must be interpreted wit caution in view of the nonindependence

the mean 4.42 measures obtained from each group (see Footnote 1).

Noteworthy features of the data base, as revealed by the distributions summarized in Tables 1, 2, and 3 include the following:

1. Table 1 shows that the source studies were primarily of behavioral methods, with systematic desensitization the most widely represented method, followed by relaxation and rehearsal/self-control/monitoring. Cognitive and minimal therapies were less widely represented, and verbal methods figured very little in the data.

2. The predominant mode of therapy was group (52% of groups), and the average therapist had some 3 years of experience, the level of an advanced postgraduate student. Therapy typically lasted for around 7 hours, and most procedures were at least moderately reproducible.

3. Table 2 shows that the most common target problems were performance anxieties, followed by physical and habit problems, social and sexual problems, and phobias.

4. Positive psychiatric diagnoses were made for only 6% of treatment groups, and

Table 2
Treatment Targets and Effect Sizes for 414 Groups

Target	*N*	%	Effect size	*SD*	*SE*
Anxiety and depression	30	7	.67	.62	.10
General anxiety	7		.16	.43	.16
Mixed phobias and anxiety	4		.95	.38	.19
Tension	9		1.10	.77	.26
Depression	10		.51	.33	.10
Phobias	76	18	1.28	.88	.10
Agoraphobia	3		2.03	.11	.06
Fear of flying	3		1.05	.90	.52
Snake phobia	40		1.23	.68	.11
Spider phobia	13		1.08	.59	.16
Rat phobia	5		2.94	1.59	.71
Dental phobia	9		.97	.71	.24
Cardiac catheterization	3		.40	.60	.35
Physical & habit problems	106	26	1.10	.85	.08
Hypertension	2		1.35	.98	.65
Headache	12		.90	.44	.12
Obesity	30		1.09	.91	.17
Underweight	2		.44	.89	.63
Heroin abuse	4		.48	.04	.02
Smoking	30		1.24	.99	.18
Nailbiting	10		1.43	.51	.16
Insomnia	12		1.40	.75	.22
Stuttering	4		−.10	.42	.21
Social and Sexual problems	76	18	.95	.75	.09
Assertion	21		1.02	.57	.12
Social inadequacy	15		.69	.58	.15
Class discussion difficulty	2		.25	.16	.11
Dating anxiety	20		1.13	.89	.20
Marital communication	4		.73	.78	.39
Sexual difficulties	14		1.02	.91	.24
Performance anxieties	126	30	.80	.71	.06
Academic underachievement	14		.69	1.48	.40
Text anxiety	70		.75	.41	.05
Public speaking anxiety	38		.95	.76	.06
Vocational indecision	4		.57	.40	.20

Note. F targets = 3.44, $p < .01$; η^2 targets = .21; F target classes = 6.38, $p < .01$; η^2 target classes = .06.

age (Mdn = 20.25 years) and educational level reflect a predominance of college students among the clients, in only 17% of whom did the target problem reach clinical severity, and who were predominantly solicited for treatment; information was rarely provided concerning concurrent medication, and most data were obtained in laboratory settings.

5. The most frequent domains of outcome measurement were fear/anxiety (39%) and social behavior (21%); most measures exhibited moderate or high specificity, and moderately low reactivity. Psychometric instruments (41%) and self-ratings (27%) were the predominant measurement techniques.

6. The typical design involved random assignment of clients (90%, with 32% involving matching) with relatively little attrition, and the same therapist or therapists in each of the treatments compared (80%). The majority of the data showed satisfactory pretreatment equivalence, and more than half (53%) of the available data indicated that the measures were single blind. The majority of the effect sizes (60%) were obtained via the preferred formula of mean difference over control standard deviation, using posttreatment scores.

Correlates of Effect Size

In interpreting the relations between coded parameters and effect size, we added careful description of the obtained distributions (see Footnote 1) to summary data in the form of the correlations presented in Table 3. Attention was focused in particular on the conditions associated with substantial deviation from the overall mean effect size of .93.

Treatment and therapist variables. The mean effect sizes for each treatment method and for the types into which they were grouped are shown in Table 1. The effect due to treatment methods was highly significant, $F(14, 339) = 3.43$, $p < .001$, and about 10% of the variance between groups was accounted for ($\eta^2 = .11$). The means associated with treatment methods ranged from .26 for study skills training to 1.52 for covert behavioral methods. The data for treatment types show mixed methods (predominantly behavioral in orientation) to yield the largest effect sizes ($M = 1.42$), followed by behavioral and cognitive methods ($M = 1.06$ and 1.00, respectively), unclassified treatments (also predominantly behavioral; $M = .78$), minimal or placebo treatments ($M = .71$), and dynamic and humanistic (verbal) therapies ($M = .40$). Among the behavioral methods, it is noteworthy that the most frequently studied method, systematic desensitization, is associated with only average outcome ($M = .97$), whereas relaxation alone appears almost as effective ($M = .90$). Although figuring in only four studies, the poor performance of study skills training is striking. Treatment mode was significantly related to effect size, $F(3, 403) = 3.91$, $p < .01$. Although individual therapy appeared the most effective mode ($M = 1.12$), it was closely followed by the predominant group mode (M = .89), and the only striking result was obtained for the rarely used couple/family mode (M = .21). Therapist experience, ranging from 0 to 8 years, was negatively correlated with effect size, $r(269) = -0.14$, $p < .01$. The distribution suggests a downward trend in mean effect size with increasing therapist experience from up to 1 year ($M = 1.17$) to 4 years (M = .62), which was reversed for the most experienced therapists ($M = .87$), although the outcome was still inferior to that obtained by neophyte therapists. However, multiple regression analysis of the data from the 27 groups for which therapist experience was recorded suggested that this was due to a tendency for studies involving less experienced therapists to be concerned with target problems yielding higher effect sizes. Entered alone, therapist experience accounted for 2% of the variance in effect size, $R^2 = .020$, $F(1, 269) = 5.61$, $p < .05$; entered after dummy variables representing 23 target problems which themselves accounted for almost 20% of the variance, $R^2 = .190$, $F(23, 247) = 2.$ $p < .01$, therapist experience no longer predicted effect size, R^2 change = .0007, $F <$ The duration of therapy also appeared negatively related to outcome until the long durations were reached, although there was no evidence of a linear relation, $r(412) =$. $p < .10$. There was no relation between reproducibility of treatment and effect size $r(412) = .03$, $p < .10$.

Client variables. The effect size for e

target, and for the five target classes into which they were grouped, are shown in Table 2. Differences between targets were highly significant, $F(29, 384) = 3.53$, $p < .001$, and accounted for over 20% of the variance in effect sizes, $\eta^2 = .21$. The means indicate large effects for phobias ($M = 1.29$) and small effects for anxiety and depression ($M = .67$), with intermediate results for physical and habit problems, ($M = 1.10$), and performance anxieties ($M = .80$). Although the grouping into target classes was inevitably somewhat arbitrary, it appears reasonably well justified by the outcomes obtained; differences between the five classes accounted for 5.9% of the variance of the scores obtained for the 414 groups. Targets yielding results discrepant from those obtained for the class to which they were assigned included mixed phobias and anxiety ($M = .95$), and tension ($M = 1.10$), cardiac catheterization stress ($M = .40$), class discussion difficulty ($M = .25$), and marital communication difficulty ($M = .73$). The class of physical and habit problems yielded divergent results for targets within it, but there was no simple differentiation between effect sizes for the somatic symptoms and for the addictive or habit disorders represented. There were no statistically dependable relations with diagnosis, education or source of clients. However, effect size was positively related to age of client, $r(412) = .17$, $p < .01$. There was no linear relation between severity/screening

Table 3
Means or Proportions on Background Variables and Their Correlations With Effect Size for 414 Groups

Variable	Mean/ proportion	*SD*	*r*	*n*
Therapy				
Mode (individual vs. others)	.41		.15**	407
Therapist experience (years)	2.91	1.93	−.14*	271
Duration (hours)	6.89	4.26	.05	396
Reproducibility	2.27	.65	.03	414
Client				
Severity/screening	1.65	.75	.04	393
Source (solicited vs. others)	.89		.05	394
Age (years)[a]	24.57	8.36	.04	393
Education (high vs. others)	.82		.00	344
Contextual				
Medication (present vs. absent)	.33		−.29*	48
Setting (lab vs. others)	.71		.08	390
Design				
Treatment *n*	11.98	7.12	−.14**	414
Control *n*	11.88	6.50	−.21**	414
Assignment of clients	2.36	.85	.05	388
N of therapists	2.14	1.31	−.06	310
Attrition of treated group (%)	10.68	13.49	−.06	1,483
Attrition of control group (%)	9.19	13.46	.00	1,486
Blindedness	2.24	.87	−.10**	1,006
Equivalence	1.26	.54	.07*	1,401
Measurement				
Domain (tractability)	.48		.09**	1,828
Reactivity	2.17	.62	.11**	1,828
Specificity	3.08	.63	.07*	1,828
Measurement technology	2.02	.90	−.11**	1,828
Follow-up (months)	.79	2.36	−.01	1,828

[a] Median age = 20.25 years.
* $p < .05$. ** $p < .01$.

and effect size, $r(390) = .04$, $p > .10$, with the largest effects ($M = 1.13$), among moderately screened clients.

Contextual variables. There was a significant effect due to setting, $F(4, 385) = 2.53$, $p < .05$. Effect sizes for home-based treatments ($M = 1.70$) were larger than for other settings. Although information on concurrent medication was not often provided, effect sizes were larger without medication ($M = 1.37$) than with it ($M = .72$), $F(1, 46) = 4.11$, $p < .05$.

Measurement variables. Each of the measurement variables was significantly related to effect size. With respect to domain, $F(7, 1820)$, $p < .001$, the most striking feature was the weak effect associated with achievement measures ($M = .28$), and personality traits ($M = .52$). There was a tendency for more specific measures to yield more favorable outcomes, $r(1826) = .07$, $p < .01$. Similarly, more favorable outcomes were associated with relatively reactive measures, $r(1826) = .11$, $p < .001$. The results for measurement technology, $F(6, 1821) = 11.09$, $p < .001$, suggested relatively favorable outcomes for self-ratings and behavioral counts, and relatively unfavorable outcomes for hard, nonpsychological data and physiological and psychometric measures. The significant effect due to length of follow-up, $F(11, 1816) = 5.22$, $p < .001$, was not accounted for by any linear relation, $r(1826) = .006$. The distribution suggests a U-function with somewhat weaker effects for intermediate lengths of follow-up ($M = .81$), than for immediate posttesting ($M = .95$), or follow-up after 4 months or more ($M = 1.04$).

Design variables. Most design variables bore at least a suggestive relation to effect size. Assignment of clients was unrelated to effect size, $F(3, 384) = 1.56$, $p > .10$. Similarly, assignment of therapists was unrelated to effect size, $F(5, 294) = 1.14$, $p > .25$. Effect size decreased with increased number of both treated clients, $r(412) = .14$, $p < .001$, and control clients, $r(412) = .21$, $p < .001$. A positive correlation was obtained between effect size and attrition in the treated group, $r(1482) = .07$, $p < .01$. The relationship between effect size and attrition in the control group, $F(5, 1480) = 2.77$, $p < .05$, was less readily interpreted, $r(1484) = .02$, $p > .25$; larger effect sizes appeared to be associated with either no attrition ($M = .96$), or a large attrition rate, with means of .91 and 1.22 for measures suffering between 30% and 40% and over 40% attrition, respectively; smaller effect sizes were associated with intermediate amounts of attrition from the control group, with means ranging from .70 to .83. Pretreatment equivalence was positively correlated with effect size, $r(1399) = .07$, $p < .05$. Blindedness of the data-gatherer was negatively correlated with effect size, $r(1004) = -.10$, $p < .01$.

Effect size calculation. There was some suggestion of a relation between type of effect size calculation and the effect size obtained, $F(3, 1824) = 1.12$, $p = .095$, and (intendedly conservative) estimates from p values yielded relatively low effect sizes ($M = .76$). The source of the effect size was modestly associated with its magnitude, $F(3, 1824) = 3.60$, $p < .01$, but this effect was largely attributable to the four effect sizes coded as other ($M = .66$). The conventional calculation of effect sizes from posttest data tended, if anything, to be associated with larger effect sizes ($M = .98$) than covariance-adjusted posttests ($M = .75$) or pre–post difference scores ($M = .80$).

Summary of correlations with effect size. Table 3 shows the correlations obtained between 23 variables and effect size. Of the variables presented in Table 3, 11 yielded significant correlations with effect size, although the significance of the smallest correlations could be spurious, in view of the nonindependence of the data obtained on each of the 414 treatment groups. In sum, these correlations suggest that effect sizes were larger for studies involving individual therapy administered by relatively inexperienced therapists and with no noncurrent medication, in which the design involved relatively small numbers of clients, who were well-matched before treatment and were assessed via specific, reactive, soft measures of tractable psychological domains that were administered nonblind. Of course, these correlations are all quite small, and zero-order correlations alone permit no conclusion as to the independence or otherwise of the effects reported.

Comparative evaluation of treatment methods. The final column of Table 1 shows the mean of the study-by-study difference

scores between each treatment and the treatments with which it was directly compared. These data indicate significant superiority of mixed, $t(23) = 2.61$, $p < .01$, and cognitive, $t(19) = 5.27$, $p < .001$, methods. There was a substantial advantage of behavioral methods as a whole ($M = .32$), $t(55) = 3.40$, $p < .01$, and significant negative scores indicated inferior results for minimal treatments, $t(35) = 6.36$, $p < .001$, dynamic and humanistic therapies, $t(12) = 2.13$, $p < .05$, and unclassified methods, $t(13) = 2.55$, $p < .05$. Elimination of minimal treatments made little difference to these comparisons, except that the inferiority of unclassified treatments became rather more marked ($M = -.38$), $t(13) = 4.47$, $p < .001$. In view of the suggestion by Smith et al. (1980, p. 119) that insight therapies are often set up as a "straw man" or quasi-control group not expected to yield strong effects, we further examined the 13 studies comparing dynamic and humanistic therapies with other treatments. In 8 studies, the dynamic or humanistic method was presented as containing no therapeutic elements unique to itself and could indeed be viewed as a straw man. In the remaining 5 studies, specific therapeutic elements were ascribed to the dynamic or humanistic therapy. The mean advantage score for the former group was −.40, and for the latter group, −.77 (rising to −.26 with the exclusion of one study yielding an extreme score of −3.06). Despite the applicability of the straw man argument to much of this data, it does not wholly account for the modest showing of dynamic and humanistic therapies.

Controlled comparisons between treatment methods. Table 4 shows the results of the 21 direct comparisons based on at least 4

Table 4
Controlled Comparisons Between Treatment Methods

Method A	Method B				
	Relaxation	Systematic desensitization	Social skills training	Mixed	Minimal
Rehearsal, self-control, and monitoring					.64*** (6)
Biofeedback	−.20 (8)			−.72* (4)	
Covert behavioral					.54* (4)
Relaxation		−.24 (13)		−.59* (5)	.29* (4)
Systematic desensitization				−.28** (7)	.50**** (15)
Reinforcement		.32 (5)			.14 (5)
Social skills training		.06 (5)			.37** (4)
Cognitive		.53**** (9)	.28 (4)		.68** (7)
Dynamic/humanistic			.35 (4)	−.93 (4)	
Unclassified	−.16 (4)	.02 (6)			.46** (6)

Note. All comparisons are Method A − Method B differences, based on 1 difference score per study and *N* of studies given in parentheses.
* $p < .10$. ** $p < .05$. *** $p < .01$. **** $p < .001$.

Table 5
Breakdown of Effect Sizes by Treatment Type and Target Class

Target class	Treatment type				
	Behavioral	Cognitive	Dynamic/ humanistic	Minimal	Total
Anxiety/depression					
ES	.74	1.34	.40	.38	.67
n	21	1	5	3	30
Phobias					
ES	1.46	.92		.66	1.28
n	56	9		11	76
Physical/habit problems					
ES	1.19	.37	.37	1.07	1.10
n	80	5	5	14	104
Social/sexual problems					
ES	1.08	1.19	.36	.55	.97
n	51	9	8	6	74
Peformance anxieties					
ES	.81	.97	.65	.36	.80
n	102	15	2	7	126
Total					
ES	1.06	.94	.40	.71	.98
n	310	39	20	41	410

Note. ES = effect size; *n* = number of groups.

studies yielding a total of at least 10 effect sizes for each of the two treatments under comparison. In assessing comparisons between purportedly active treatments and minimal or placebo methods, a directional, one-tailed test was adopted, in view of the a priori expectation that active treatments would be more effective. Using this criterion, six methods were significantly superior to minimal or placebo treatments. These were rehearsal/self-control, $t(5) = 3.54$, $p < .01$; covert behavioral, $t(3) = 2.63$, $p < .05$; systematic desensitization, $t(14) = 4.71$, $p < .001$; social skills training, $t(3) = 3.20$, $p < .05$; cognitive therapy, $t(6) = 3.11$, $p < .05$; and unclassified treatments, $t(5) = 2.81$, $p < .05$. Marginally significant was the superiority of relaxation over minimal treatments, $t(3) = 2.32$, $p < .10$. All comparisons between active treatments were evaluated via two-tailed tests. Cognitive therapy was superior to systematic desensitization, $t(8) = 2.86$, $p < .05$. Mixed methods also were superior to systematic desensitization, $t(6) = 2.86$, $p < .05$. Mixed methods also appeared superior to relaxation, $t(4) = 2.37$, $p < .10$, and to biofeedback, $t(3) = 2.45$, $p < .10$, but these low-power comparisons were only marginally significant. No other comparisons approached statistical significance. Of particular interest because based on a fair number of studies were the apparent equivalence of biofeedback and relaxation, $t(7) = .74$; and of systematic desensitization and relaxation $t(12) = .94$.

Treatment × Client Interaction

Table 5 shows the breakdown of the mean effect sizes obtained for 410 groups by treatment type and target class. Excluded from this analysis were the 4 groups classified as of mixed type, because they combined major elements of more than one of the coded types. A regression-model analysis of variance (ANOVA) of these data yielded a main effect due to treatment type, $F(3, 391) = 5.5$ $p < .001$. There was no significant effect due to target class ($F < 1$), or to the two-way interaction, $F(11, 391) = 1.41$, $p < .10$. The

only suggestive evidence of outcomes discrepant from that predicted from the marginal totals occurred in relation to physical and habit problems, which appear effectively treated by minimal treatments (M = 1.07) but ineffectively treated by cognitive methods (M = .37).

Table 6 shows a similar breakdown by treatment type and domain of outcome measurement. In addition to main effects corresponding to those reported earlier, ANOVA revealed a significant interaction, $F(21, 1786) = 1.77$, $p < .02$. Although based on relatively few effect sizes, noteworthy cell means in Table 6 include the relatively poor performance of cognitive methods with adjustment measures (M = .28) and emotional/somatic disorder (M = .36); the fair showing of dynamic and humanistic methods with physiological stress measures (M = .86); and the relatively good results of cognitive therapy (M = 1.39), and poor results of minimal treatments (M = .06), with achievement measures. Inevitably, the results of the breakdowns of Tables 5 and 6 are dominated by the preponderance of behavioral methods in the data.

Multiple Regression Analysis

Dummy variables were created to represent 14 treatment methods, 3 treatment

Table 6
Breakdown of Effect Sizes by Treatment Type and Domain of Measurement

Domain	Treatment type				
	Behavioral	Cognitive	Dynamic/ humanistic	Minimal	Total
Fear/anxiety					
ES	1.14	.93	.45	.72	1.06
n	558	78	11	72	719
Self-esteem					
ES	1.19	1.24	.35	.30	.95
n	103	14	31	16	164
Adjustment					
ES	1.50	.28	.09	.16	.96
n	39	11	8	8	66
Achievement					
ES	.26	1.39	.45	−.06	.28
n	118	6	3	12	139
Personality traits					
ES	.53	1.03	.89	.32	.52
n	23	1	1	5	30
Social behavior					
ES	.99	1.07	.31	.62	.95
n	264	87	13	31	391
Emotional/somatic disorder					
ES	.97	−.36	.18	.87	.89
n	170	5	10	20	205
Physiological stress					
ES	1.02	1.63	.86	.44	.97
n	82	5	4	13	104
Total					
ES	1.10	.97	.35	.57	.93
n	1357	203	81	177	1818

Note. ES = effect size; *n* = number of effect sizes.

Table 7
Impact of Five Sets of Variables on Effect Size

Set	1,802 measures				407 groups			
	Alone		With all sets		Alone		With all sets	
	R^2	F	ΔR^2	F	R^2	F	ΔR^2	F
Methods	.064	8.67***	.049	7.85***	.112	3.54***	.089	3.67***
Targets	.111	7.66***	.076	5.88**	.217	3.60***	.091	1.77**
Sample size	.046	43.47***	.034	38.11***	.068	14.78***	.049	13.82***
Mode	.007	13.36***	.005	11.21***	.021	8.70**	.009	5.08*
Measurement variables	.043	20.04***	.037	20.74***				

* $p < .05$. ** $p < .01$. *** $p < .001$.

types, and 29 target problems. Ideally, a multiple regression analysis would have included alongside these dummy variables all other variables showing evidence of association with effect size. Unfortunately, missing values arising from incomplete reporting by source study authors reduced the number of variables on which sufficient data were available to permit a dependable analysis. In addition to these sets of dummy variables, complete data were available for 1,802 effect sizes, representing 407 treatment groups, with respect to treatment mode; treatment and control group *n*s; and the four measurement variables, domain, reactivity, specificity, and measurement technology.

The rightmost section of Table 7 shows the variance in mean effect size for each treatment group accounted for by four sets of variables and presents R^2 values for each set of variables alone, together with the change in R^2 associated with the addition of that set at the last step in the regression equation. Taken together, the four sets of variables accounted for over 35% of the variance (R^2 = .362), $F(46, 360) = 4.45$, $p < .001$. As shown in Table 7, treatment methods accounted for 11% of the variance when entered alone (R^2 = .112), $F(14, 392) = 3.54$, $p < .001$, and this effect was only slightly diminished by taking account of the other sets of variables, R^2 change = .089, $F(14, 360) = 3.67$, $p < .001$. In contrast, the initially larger impact of the 29 treatment targets (R^2 = .217), $F(29, 377) = 3.60$, $p < .001$, was reduced to a level comparable to that of the treatment methods (R^2 change = .091), $F(29, 360) = 1.77$, $p < .01$, when the other sets of variables were taken into account. On the left in Table 7 are presented the results of an analysis of the 1,802 effect size measures, in which four measurement variables were also included. Overall, only 22% of the variance was accounted for by the five sets of variables, (R^2 = .219), $F(50, 1751) = 9.79$, $p < .001$. Entered alone, the 14 methods accounted for abou 6% of the variance (R^2 = .64), $F(14, 1787)$ = 8.67, $p < .001$, which was reduced to 5% (R change = .049), $F(14, 1751) = 7.85$, $p < .001$ when the other four sets of variables wer taken into account. Entered alone, the 2 targets accounted for 11% of the varianc (R^2 = .11), $F(29, 1772) = 7.66$, $p < .001$; thi was reduced to about 8% (R^2 change = .76 $F(29, 1751) = 5.88$, $p < .001$, when the othe 4 sets of variables were included in the equ tion. Overall, the results presented in Tab 7 indicate that the effects of each set of var ables were largely, although not wholly, i dependent of those of the other sets of va ables. Estimates of the impact of treatme methods ranged from 5% to 11% of the va ance. Each of the four variable sets includ in both analyses accounted for less varian in the analysis of individual measures, whe more variance remained unaccounted f (78% as opposed to 64%) despite the int duction of the four measurement variable

Standardized regression coefficients each variable in the analysis of 1,802 eff size measures are shown in Table 8. On left in the table are presented the betas tained for each variable when the set of v ables to which it was assigned was ente

alone; on the right are the betas obtained when all variables were in the equation. The close correspondence between the two columns of figures, consistent with the high degree of independence between variable sets noted in Table 7, indicates that controlling for other variable sets has little impact on the relation between each variable and effect size. Among the dummy variables representing treatment methods, a possible exception to this trend was obtained for flooding, whose significant beta of .085, $F(1, 1751) = 10.41$, $p < .01$, fell to $-.005$ ($F < 1$) when the other variable sets were entered into the equation. It is noteworthy that the direction of impact of treatment n was reversed by taking account of control n in the variable set comprising these 2 variables and that the impact of reactivity was reduced from a zero-order r of .11 to a beta of .06 when entered alongside the other 3 measurement variables and to a beta of $-.013$ when all variable sets were included. In these latter circumstances, specificity and technology were the most powerful measurement variables, with betas of .103, $F(1, 1751) = 14.49$, $p < .001$, and $-.121$, $F(1, 1751) = 17.66$, $p < .001$, respectively. Four variables yielding significant correlations with effect size in Table 3 were excluded from the multiple regression analysis on account of missing values; of the 7 variables with such significant correlations that did appear in the multiple regression equations, 5 yielded significant betas of unchanged sign when all variable sets were included in the equation. These were specificity, technology, domain ($\beta = -.062$), $F(1, 1751) = 5.03$, $p < .05$, control n ($\beta = -.335$), $F(1, 1751) = 64.85$, $p < .001$, and mode ($\beta = .086$), $F(1, 1751) = 9.74$, $p < .01$.

Standardized regression coefficients corresponding to those shown in Table 8 were also obtained for all but the measurement variables in the analysis of the mean effect sizes obtained for each of 407 treatment groups. Consistent with the multiple R^2 data

Table 8
Standardized Regression Coefficients for Treatment, Design, and Measurement Variables

Variable	Set alone		With all sets	
	β	F	β	F
Rehearsal, self-control, & monitoring	.064	4.93*	.052	3.13-
Biofeedback	.021	<1	.000	<1
Covert behavioral	.149	29.98***	.112	17.28***
Flooding	.085	10.41**	−.005	<1
Relaxation	.055	3.30	.052	2.77
Systematic desensitization	.111	9.37**	.126	12.36***
Reinforcement	.053	3.88*	.052	3.30
Modeling	.163	35.62***	.142	24.60***
Social skills training	.040	2.14	.065	5.13*
Study skills training	−.031	1.64	.011	<1
Cognitive	.113	12.94***	.160	26.36***
Dynamic/humanistic	−.042	2.39	.001	<1
Mixed	.207	45.17***	.202	45.63***
Unclassified	.026	<1	−.010	<1
Behavioral	.162	22.46***	.159	22.96***
Verbal	−.040	2.16	−.004	<1
Cognitive	.110	11.88***	.164	27.08***
Mode	.086	13.36***	.086	9.74**
Treatment N	.272	50.94***	.322	63.71***
Control N	−.355	86.71***	−.335	64.85***
Domain	−.125	26.02***	−.062	5.03**
Reactivity	−.064	5.70	−.013	<1
Specificity	.163	40.67	.103	14.49
Technology	−.111	16.24***	−.121	17.66***

Note. Variables representing the 3 therapy types, behavioral, verbal, and cognitive, were entered in place of the 14 methods in a separate analysis.
* $p < .05$. ** $p < .01$. *** $p < .001$.

of Table 7, these betas were larger than those shown in Table 8. Their relative magnitudes were generally very similar, however. The only exception was obtained for the dummy variable representing cognitive therapy, whose beta rose from a nonsignificant .099, $F(1, 392) = 2.51$, to .216, $F(1, 360) = 12.56$, $p < .001$, when the other three variable sets were included in the equation. This suggests that cognitive therapy was typically studied under somewhat adverse conditions with respect to the other sets of variables, yielding more favorable outcomes with statistical control for this adversity.

Discussion

The findings of the present study are broadly in agreement with those of most previous reviews, whether conventional or meta-analytic. The evidence from the relatively well-conducted source studies conforming to the design in which two or more treatment groups are compared with a control group is consistent with the less restrictive but largely independent survey by Smith et al. (1980) in suggesting a mean effect size approaching one standard deviation. The effects of different treatment methods were not, on the whole, impressively different from one another, accounting for less variance than differences between target problems under treatment. Simultaneous controlled comparisons among the behavioral methods with which most of our data were concerned suggested that these have broadly similar effects. The impact of client, therapist, design, and measurement variables was generally modest, and although treatment targets and design features appeared to account for a large proportion of the variance in effect sizes, some of this variance was attributable to covariation with other study characteristics, such that the overall percentages of variance accountable for in our two regression analyses were 36% and 22%. There was little convincing evidence of differential impact of treatment methods with clients with differing problems or with outcome measurements in different psychological domains. Thus, the broad thrust of our findings suggests moderate, generally uniform, effects of therapy, with a large proportion of the variance in effect sizes unaccounted for.

On the other hand, some clear exceptions to these general trends are worthy of mention. The present data revealed a modest but undeniable superiority of behavioral and cognitive methods and a corresponding relative inferiority of dynamic and humanistic (verbal) methods, in the results of simultaneous controlled comparisons; statistical control of variation in other study characteristics via multiple regression analysis further highlighted the apparent superiority of cognitive therapy and did little to diminish the impact of the treatment methods generally.

Specific comparisons between treatment methods present together in at least four studies yielded some quite clearcut results. First, active treatments were superior to minimal, placebo treatments in the majority of such comparisons. Of more interest were the comparisons among active treatments. In view of the procedural and theoretical commonalities among relaxation, biofeedback and systematic desensitization, it is noteworthy that these three treatments figured prominently among those behavioral methods that come out relatively unfavorably in the data of Table 4. In view of recent controversy (Ledwidge, 1978, 1979; Mahoney & Kazdin, 1979), the advantage of cognitive therapy over desensitization is of particular note. The nine studies making this comparison were primarily concerned with performance anxiety (6 studies); all were with student clients and none involved depression. None was included in Ledwidge's (1978, Table 1) review of earlier studies. The authors of all 9 studies interpreted their outcomes as favoring cognitive therapy over desensitization, except that in one study a combined treatment (coded by our conventions as cognitive) was superior to either component in isolation. Thus, the conclusions of meta-analysis were congruent with those drawn by individual authors.

As with all our findings, these results cannot be interpreted without reference to the nature of the data base. They should not be taken to imply that cognitive therapy is superior to orthodox behavioral methods in the treatment of depression or with clinical populations. Their major value is in their demonstration that meta-analytic methods can yield clearcut findings within a closely de

fined subset of a larger data base and that the overall picture of broadly comparable treatment outcomes does not rule out the existence of tangible exceptions to the general rule.

The nature of our data base is itself an issue warranting comment. Overwhelmingly, studies simultaneously comparing two or more treatments with a control group are concerned with behavioral methods, applied in what can only be described as an analogue context, in which students are recruited for brief treatment of relatively circumscribed and minimally disabling target problems, and outcomes are assessed immediately posttherapy. Few indeed are the studies of dynamic therapy; almost as rare is the involvement of clinically referred clients in controlled outcome studies of the kind reviewed here. Similar patterns were reported by Smith et al. (1980), despite the less exacting design requirements used in the selection of their largely independent data base. No meta-analysis can transcend the limitations of the data gathered by researchers undertaking its source studies. The frequency data reported in Tables 1 to 3 show how unrepresentative of clinical practice the literature reviewed is. This concentration of outcome research at the analogue end of the clinical–analogue continuum places severe limitations on its generalizability to everyday practice and its utility in the formulation of policy concerning service provision and the orientation of training and professional development (Vandenbos, 1980; Shapiro & Shapiro, in press).

Our overall effect size of .93 was slightly larger than the .68 and .85 obtained by Smith and Glass (1977) and by Smith et al. (1980), respectively. A possible explanation of this slight difference lies in our exclusion of dissertations. Smith et al. (1980) found a mean effect size for dissertations of .66, whereas that for journal articles was .87. Glass et al. (1981, p. 66) present comparisons from other meta-analytic investigations indicating similar discrepancies between journal article and dissertation data, with journal article data yielding more favorable results. Since Smith et al. (1980) found that the rated internal validity of dissertation studies was not inferior to that of journal articles, their omission from the present meta-analysis may have resulted in an unwarranted inflation of the overall effect size estimate. Thus the gain in replicability achieved by the elimination of dissertation studies, in accordance with the preference of such critics as Rachman and Wilson (1980), may be offset by the introduction of bias, possibly reflecting a prejudice against the null hypothesis (Greenwald, 1975) in published work (Shapiro & Shapiro, 1982).

The sparse representation and poor showing of dynamic and humanistic methods in our data are striking. The professional attitudes and values of dynamically oriented therapists may well militate against their conducting and publishing contrast-group outcome studies of the kind reviewed here. Adherents of these approaches may consider them ill-served by studies in which they were predominantly represented by group methods (80% of the total effect sizes for these therapies) that were administered by therapists with a mean 3.7 years of experience, for a mean duration of 11 hours, with 75% of the data obtained immediately after treatment and with only 33% of the data obtained in clinical settings. On the other hand, the target problems predominantly addressed by these methods (51% social and sexual problems, 17% anxiety and depression, and 16% physical and habit problems) appear appropriate enough, and the mean age of 27 years for clients receiving these treatments indicates that a substantial proportion were not college students. The majority of the few instances of dynamic and humanistic therapy were interpretable as straw man treatment groups, not expected to yield good results by the investigators. Smith et al. (1980) found that the apparent allegiance of the researcher was a good predictor of the outcome obtained for a given treatment. On the other hand, whether or not dynamic or humanistic treatment groups appearing in the present meta-analysis were straw men, their apparent effects were consistently modest in comparison with other treatments. Nonetheless, the volume and representativeness of these data are insufficient to warrent their interpretation as supporting an adverse evaluation of the verbal therapies.

The major impact of target problem on effect size is consistent with clinicians' recognition that some problems are more readily overcome than others, irrespective of the

treatment approach. Specifically, the large effects for simple phobias, in contrast to the more modest effects for anxiety and depression, are consistent with the view that research focusing on the former targets may yield overoptimistic results, which may not generalize to the latter targets. Similarly, the modest superiority of individual over group therapy accords with clinical lore; in this case, however, the implication is that the emphasis in outcome research on group methods may underestimate treatment outcomes. Equally consistent with clinical lore is the greater impact of therapy on measures of anxiety and self-esteem, and on relatively soft and treatment-specific measures. Our explanation of the apparent superiority of outcomes with relatively inexperienced therapists in terms of the tendency for such therapists to be employed in studies of relatively tractable target problems is also consistent with the view of therapy research held by many clinicians.

Our analysis of design features yielded generally reassuring results. If anything, client attrition and poor matching of groups were associated with weaker rather than stronger treatment effects, suggesting that these design deficiencies are unlikely to inflate the estimated impact of treatment. The negative relation between effect size and sample size may well reflect the tendency for studies to be written up and published only when statistically significant results are obtained, such that publication of data from small samples necessitates more powerful effects. Other design features appeared unrelated to effect size apart from the disconcerting tendency for larger effects to emerge from nonblind assessments. This underlines the importance of reducing this source of bias in future research.

Our efforts to refine the meta-analytic methods applied by Smith et al. (1980) have yielded tangible benefits. First, the focus on studies including two or more active treatments produced a data base considerably richer in potential for direct comparisons free of bias arising from noncomparable studies of the different methods; in consequence, we are able to conclude rather more firmly that treatment methods make a difference, albeit modest, to effect size. Second, our refinement of the coding system with respect to methods of assessment yielded more precise conclu sions concerning their impact on effect siz Although replicating Smith et al.'s (198(finding that reactivity (here differentiate into specificity and other aspects of reacti ity) is a clear correlate of effect size, our da suggested that this effect is primarily attril utable to the specificity component of Smit et al.'s (1980) reactivity dimension. Furthe more, this effect was equaled by a statistical independent effect associated with the sof ness of the measurement technology use which was a new variable devised for th present analysis.

We do not claim that a meta-analysis suc as the one reported here can alone furni definitive answers to the perennially vexi and complex questions concerning the in pact of psychological treatments. The int gration of data from diverse studies is fraug with difficulties, and meta-analysis is n more than an attempt to apply to that ta the judgmental, analytic, and computation skills of the empirical researcher. As in a empirical research, however, the outcom obtained are a function of the assumptio resources, and skills applied by the inves gator. Nonetheless, the present study poi clearly to three broad conclusions: First, t average impact of psychological treatmer in recently published comparative outcor research approaches one standard deviati unit; second, the modest differences betwe treatment methods are largely independe of other factors influencing outcome; a third, contemporary outcome research is n representative of clinical practice.

References

Arkin, R., Cooper, H., & Kolditz, T. A statistical rev of the literature concerning the self-serving bias interpersonal influence situations. *Journal of Pers ality*, 1980, *48*, 435–448.

Bandura, A. Self-efficacy: Toward a unifying theor behavioral change. *Psychological Review*, 1977, 191–215.

Barber, T. X. Commentary on Rosenthal and Rul *Behavioral and Brain Sciences*, 1978, 388–390.

Beck, A. T. *Cognitive therapy and the emotional orders*. New York: International Universities Pr 1976.

Cautela, J. Covert sensitization. *Psychological Rep* 1967, *20*, 459–468.

Cook, T. D., & Leviton, L. C. Reviewing the literat A comparison of traditional methods with meta-a ysis. *Journal of Personality*, 1980, *48*, 449–472.

Cooper, H. M. Statistically combining independent s

ies: A meta-analysis of sex differences in conformity research. *Journal of Personality and Social Psychology*, 1979, *37*, 131–135.

D'Zurilla, T. J., & Goldfried, M. R. Problem solving and behavior modification. *Journal of Abnormal Psychology*, 1971, *78*, 107–126.

Eagly, A. H., & Carli, L. L. Sex of researchers and sex-typed communications as determinants of sex differences in influencibility: A meta-analysis of social influence studies. *Psychological Bulletin*, 1981, *90*, 1–20.

Elashoff, J. D. Commentary on Rosenthal & Rubin. *Behavioral and Brain Sciences*, 1978, *3*, 392.

Ellis, A. *Reason and emotion in psychotherapy*. New York: Lyle Stuart, 1962.

Eysenck, H. J. An exercise in mega-silliness. *American Psychologist*, 1978, *33*, 517.

Finney, D. J. *Probit analysis* (3rd Edn). Cambridge, England: Cambridge University Press, 1972.

Frank, J. D. The present status of ourcome studies. *Journal of Consulting and Clinical Psychology*, 1979, *47*, 310–316.

Franks, C. M., & Wilson, G. T. *Annual review of behavior therapy: Theory and practice* (Vol. 6). New York: Brunner/Mazel, 1978.

Gallo, P. S. Meta-analysis—A mixed meta-phor. *American Psychologist*, 1978, *33*, 515–517.

Glass, G. V. Integrating findings: The meta-analysis of research. *Review of Research in Education*, 1978, *5*, 351–379.

Glass, G. V., McGaw, B., & Smith, M. L. *Meta-analysis in social research*. Beverely Hills, Calif.: Sage Publications, 1981.

Glass, G. V., & Smith, M. L. The effects of class size on achievement. *Journal of Education and Policy Studies*, 1979, *1*, 2–16.

Goldfried, M. R. Systematic desensitization as training in self-control. *Journal of Consulting and Clinical Psychology*, 1971, *37*, 228–234.

Greenwald, A. G. Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 1975, *82*, 1–20.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. Effects of cooperative, competitive and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, 1981, *89*, 47–62.

Kanfer, F. H., & Karoly, P. Self-control: A behavioristic excursion into the lion's den. *Behavior Therapy*, 1972, *3*, 398–416.

Kazdin, A. E. Evaluating the generality of findings in analogue therapy research. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 673–686.

Kazdin, A. E., & Wilson, G. T. *Evaluation of behavior therapy: Issues, evidence and research strategies*. Cambridge, Mass.: Ballinger, 1978.

Ledwidge, B. Cognitive behavior modification: A step in the wrong direction? *Psychological Bulletin*, 1978, *85*, 353–375.

Ledwidge, B. Cognitive behavior modification or new ways to change minds: Reply to Mahoney and Kazdin. *Psychological Bulletin*, 1979, *86*, 1050–1053.

Luborsky, L., Singer, B., & Luborsky, L. Comparative studies of psychotherapies. *Archives of General Psychiatry*, 1975, *32*, 995–1008.

Mahoney, M. J., & Kazdin, A. E. Cognitive behavior modification: Misconceptions and premature evacuation. *Psychological Bulletin*, 1979, *86*, 1044–1049.

Mansfield, R. S., & Bussey, T. V. Meta-analysis of Research: A rejoinder to Glass. *Educational Researcher*, 1977, *6*, 3.

Mathews, A. Fear-reduction research and clinical phobias. *Psychological Bulletin*, 1978, *85*, 390–404.

McFall, R., & Twentyman, C. Four experiments on the relative contribution of rehearsal, modeling and coaching to assertion training. *Journal of Abnormal Psychology*, 1973, *81*, 199–218.

Meichenbaum, D. B. *Cognitive behavior modification: An integrative approach*. New York: Plenum, 1977.

Meltzoff, J., & Kornreich, M. *Research in psychotherapy*. New York: Atherton, 1970.

Nie, N. H, Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. *Statistical Package for the Social Sciences* (2nd ed.). New York: McGraw-Hill, 1975.

Rachman, S. J., & Hodgson, R. *Obsessions and compulsions*. Englewood Cliffs, N.J.: Prentice Hall, 1980.

Rachman, S. J., & Wilson, G. T. *The effects of psychological therapy: Second enlarged edition*. New York: Pergamon Press, 1980.

Rosenthal, R., & Rubin, D. B. Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, 1978, *3*, 377–415.

Shapiro, D. A. Science and psychotherapy: The state of the art. *British Journal of Medical Psychology*, 1980, *53*, 1–10.

Shapiro, D. A., & Shapiro, D. The 'double standard' in evaluation of psychotherapies. *Bulletin of the British Psychological Society*, 1977, *30*, 209–210.

Shapiro, D. A., & Shapiro, D. Comparative therapy outcome research: Methodological implications of meta-analysis. *Journal of Consulting and Clinical Psychology*, in press.

Shapiro, D. A., & Shapiro, D. Meta-analysis of comparative therapy outcome research: A critical appraisal. *Behavioural Psychotherapy*, 1982, *10*, 4–25.

Sloane, R. B., Staples, F. R., Cristol, A. H., Yorkston, N. J., & Whipple, K. *Psychotherapy versus behavior therapy*. Cambridge, Mass.: Harvard University Press, 1975.

Smith, M. L. Sex bias in counseling and psychotherapy. *Psychological Bulletin*, 1980, *87*, 392–407.

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

Smith, M. L., Glass, G. V., & Miller, T. I. *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press, 1980.

Strube, M. J., & Garcia, J. E. A meta-analytic investigation of Fiedler's contingency model of leadership effectiveness. *Psychological Bulletin*, 1981, *90*, 307–321.

Strube, M. J., & Hartmann, D. P. A critical appraisal of meta-analysis. *British Journal of Clinical Psychology*, 1982, *21*, 129–139.

Vandenbos, G. R. *Psychotherapy: Practice, research, policy*. Beverly Hills, Calif.: Sage Publications, 1980.

White, K. R. The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 1982, *91*, 461–481.

(Appendix follows on next page)

Appendix

Studies Included in the Meta-Analysis

Each reference is followed by the number (or numbers) assigned to the study (or studies) obtained from it as listed in Table A of the supplementary documentation available from JSAS (see Footnote 1).

Adesso, V. J., Vargas, J. M., & Siddall, J. W. The role of awareness in reducing nail-biting behavior. *Behavior Therapy*, 1979, *10*, 148–154. (1)

Aiduk, R., & Karoly, P. Self-regulatory techniques in the modification of non-assertive behavior. *Psychological Reports*, 1975, *36*, 895–905. (2)

Alden, L., Safran, J., & Weideman, R. A comparison of cognitive and skills training strategies in the treatment of unassertive clients. *Behavior Therapy*, 1978, *9*, 843–846. (3)

Anton, W. D. An evaluation of outcome variables in the systematic desensitisation of test anxiety. *Behavior Research and Therapy*, 1976, *14*, 217–224. (4)

Argyle, M., Bryant, B., & Trower, P. Social skills training and psychotherapy: A comparative study. *Psychological Medicine*, 1974, *4*, 435–443. (5)

Asirdas, S., & Beech, H. R. The behavioral treatment of sexual inadequacy. *Journal of Psychosomatic Research*, 1975, *19*, 345–353. (6)

Bander, K. W., Steinke, G. V., Allen, G. J., & Mosher, D. L. Evaluation of three dating-specific treatment approaches for heterosexual dating anxiety. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 259–265. (7)

Barbarin, O. A. Comparison of symbolic and overt aversion in the self-control of smoking. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 1569–1571. (8)

Barkley, R. A., Hastings, J. E., & Jackson, T. L. The effects of rapid smoking and hypnosis in the treatment of smoking behavior. *The International Journal of Clinical and Experimental Hypnosis*, 1977, *25*, 7–17. (9)

Barrett, C. J. Effectiveness of widows' groups in facilitating change. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 20–31. (10)

Bedell, J. R. Systematic desensitisation, relaxation-training and suggestion in the treatment of test anxiety. *Behavior Research and Therapy*, 1976, *14*, 309–311. (11)

Berlin, J. S., & Dies, R. R. Differential group structure: The effects on socially isolated college students. *Small Group Behavior*, 1974, *5*, 462–472. (12)

Blanchard, E. B., Theobald, D. E., Williamson, D. A., Silver, B. V., & Brown, D. A. Temperature biofeedback in the treatment of migraine headaches. *Archives of General Psychiatry*, 1978, *35*, 581–588. (13)

Borkovec, T. D. Heart-rate process during systematic desensitisation and implosive therapy for analog anxiety. *Behavior Therapy*, 1974, *5*, 636–641. (14)

Branham, L., & Katahn, M. Effectiveness of automated desensitisation with normal volunteers and phobic patients. *Canadian Journal of Behavioral Science*, 1974, *6*, 234–245. (15, 16)

Cautela, J. R., Flannery, R. B., & Hanley, S. Covert modeling: An experimental test. *Behavior Therapy*, 1974, *5*, 494–502. (17)

Chang-Liang, R., & Denney, D. R. Applied relaxation as training in self-control. *Journal of Counseling Psychology*, 1976, *23*, 183–189. (18)

Chesney, M. A., & Shelton, J. L. A comparison of muscle relaxation and electromyogram biofeedback treatments for muscle contraction headache. *Journal of Behavior Therapy and Experimental Psychiatry*, 1976, *7*, 221–225. (19)

Cox, D. J., Freundlich, A., & Meyer, R. G. Differential effectiveness of electromyograph feedback, verbal relaxation instructions, and medication placebo with tension headaches. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 892–898. (20)

Curran, J. P. Social skills training and systematic desensitization in reducing dating anxiety. *Behavior Research and Therapy*, 1975, *13*, 69–77. (21)

Curran, J. P., & Gilbert, F. S. A test of the relative effectiveness of a systematic desensitisation program and an interpersonal skills training program with date anxious subjects. *Behavior Therapy*, 1975, *6*, 510–521. (22)

Deffenbacher, J. L., Mathis, H., & Michaels, A. O. Two self-control procedures in the reduction of targeted and nontargeted anxieties. *Journal of Counseling Psychology*, 1979, *26*, 120–127. (23)

Deffenbacher, J. L., & Parks, D. H. A comparison of traditional and self-control desensitisation. *Journal of Consulting and Clinical Psychology*, 1979, *26*, 93–97. (24)

Deffenbacher, J. L., & Payne, D. M. J. Two procedures for relaxation as self-control in the treatment of communication apprehension. *Journal of Consulting and Clinical Psychology*, 1977, *24*, 255–258. (25)

Denney, D. R. Active, passive and vicarious desensitisation. *Journal of Counseling Psychology*, 1974, *21*, 369–375. (26)

Denney, D. R., & Rupert, P. A. Desensitisation and self-control in the treatment of text anxiety. *Journal of Counseling Psychology*, 1977, *24*, 272–280. (27)

Denney, D. R., & Sullivan, B. J. Desensitisation and modeling treatments of spider fear using two types of scenes. *Journal of Consulting and Clinical Psychology*, 1976, *44*, 573–579. (28)

Diament, C., & Wilson, T. G. An experimental investigation of the effects of covert sensitization in an analogue eating situation. *Behavior Therapy*, 1975, *4*, 499–509. (29)

Eliott, C. H., & Denney, D. R. Weight control through covert sensitisation and false feedback. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 842–850. (30)

Emmelkamp, P. M. G. Self observation versus flooding in the treatment of agoraphobia. *Behavior Research and Therapy*, 1974, *12*, 229–237. (31)

Epstein, N., & Jackson, E. An outcome study of short-term communication training with married couples. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 207–212. (32)

Finger, R., & Galassi, J. P. Effects of modifying cognitive versus emotionality responses in the treatment of test anxiety. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 280–287. (33)

Foa, E. B., Blau, J. S., Prout, M., & Latimer, P. Is horror a necessary component of flooding (implosion)? *Behavior Research and Therapy*, 1977, *15*, 397–402. (34)

Fuchs, C. Z., & Rehm, L. P. A self-control behavior therapy program for depression. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 206–215. (35)

Fremouw, W. J., & Zitter, R. E. A comparison of skills training and cognitive restructuring–relaxation for the treatment of speech anxiety. *Behaviour Therapy*, 1978, *9*, 248–259. (36)

Frohwirth, R. A., & Foreyt, T. P. Aversive conditioning treatment of overweight. *Behavior Therapy*, 1978, *9*, 861–872. (37)

Gatchel, R. J., Hatch, J. P., Watson, P. J., Smith, D., & Gaas, E. Comparative effectiveness of voluntary heart rate control and muscular relaxation as active coping skills for reducing speech anxiety. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 1093–1100. (38)

Girodo, M., & Roehl, J. Cognitive preparation and coping self-talk: Anxiety management during the stress of flying. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 978–989. (39)

Glass, C. R., Gottman, J. M., & Shmurak, S. H. Response-acquisition and cognitive self-statement modification approaches to eating-skills training. *Journal of Counseling Psychology*, 1976, *23*, 520–526. (40)

Goldfried, M. R., Lineham, M. M., & Smith, J. L. Reduction of test anxiety through cognitive restructuring. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 32–39. (41)

Gulanick, N. G., Woodburn, L. T., & Rimm, D. C. Weight gain through self-control procedures. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 536–539. (42)

Hall, S. M., Hall, R. G., Deboer, G., & O'Kulitch, P. Self and external management compared with psychotherapy in the control of obesity. *Behavior Research and therapy*, 1977, *15*, 89–95. (43)

Hannie, T. J., & Adams, H. E. Modification of agitated depression by flooding: A preliminary study. *Journal of Behavior Therapy and Experimental Psychiatry*, 1974, *5*, 161–166. (44)

Haynes, S. N., Griffin, P. Mooney, D., & Parise, M. Electromyographic biofeedback and relaxation instructions in the treatment of muscle contraction headaches. *Behavior Therapy*, 1975, *6*, 672–678. (45)

Holroyd, K. A. Cognition and desensitization in the group treatment of test anxiety. *Journal of Consulting and Clinical Psychology*, 1976, *44*, 991–1001. (46)

Holroyd, K. A., & Andrasik, F. Coping and the self-control of chronic tension headache. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 1036–1045. (47)

Horne, A. M., & Matson, J. L. A comparison of modeling, desensitization, flooding, study skills and control groups for reducing test anxiety. *Behavior Therapy*, 1977, *8*, 1–8. (48)

Hurley, A. Covert reinforcement: The contribution of the reinforcing stimulus to treatment outcome. *Behavior Therapy*, 1976, *7*, 374–378. (49)

Israel, E., & Beiman, I. Live versus recorded relaxation training: A controlled investigation. *Behavior Therapy*, 1977, *8*, 251–254. (50)

Jaffe, P. G., & Carlson, P. M. Relative efficacy of modeling and instructions in eliciting social behavior from chronic psychiatric patients. *Journal of Consulting and Clinical Psychology*, 1976, *44*, 200–207. (51)

Jeffrey, D. B., & Christensen, E. R. Behavior therapy versus "will power" in the management of obesity. *Journal of Psychology*, 1975, *90*, 303–311. (52)

Kantorowitz, D. A., Walters, J., & Pezdek, K. Positive versus negative self-monitoring in the self-control of smoking. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 1148–1150. (53)

Kaplan, R. M., McCordick, S. M., & Twitchell, M. Is it the cognitive or the behavioral component which makes cognitive-behavior modification effective in text anxiety? *Journal of Counseling Psychology*, 1979, *26*, 371–377. (54)

Kass, W., & Gilner, F. H. Drive level, incentive conditions, and systematic desensitization. *Behavior Research and Therapy*, 1974, *12*, 99–106. (55)

Katz, R. C., Thomas, S. L., & Williamson, P. Effects of self-monitoring as a function of its expected benefits and incompatible response training. *Psychological Record*, 1976, *26*, 533–540. (56)

Kelly, A. H., & Curran, J. P. Comparison of a self-control approach and an emotional coping approach to the treatment of obesity. *Journal of Consulting and Clinical Psychology*, 1976, *44*, 683. (57)

Kendall, P. C., Williams, L., Pechacek, T. F., Graham, L. E., Shisslak, C. & Herzoff, N. Cognitive-behavioral and patient education interventions in cardiac catheterization procedures: The Palo Alto medical psychology project. *Journal of Consulting and Clinical Psychology*, 1979, *47*, 49–58. (58)

Kessler, S. Treatment of overweight. *Journal of Counseling Psychology*, 1974, *21*, 395–398. (59)

Kilmann, P. R. Marathon group therapy with female narcotic addicts. *Psychotherapy Theory Research and Practice*, 1974, *11*, 339–342. (60)

Kipper, D. A., & Giladi, D. Effectiveness of structured psychodrama and systematic desensitization in reducing test anxiety. *Journal of Counseling Psychology*, 1978, *25*, 499–505. (61)

Kirsch, I., & Henry, D. Extinction versus credibility in the desensitization of speech anxiety. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 1052–1059. (62)

Kirsch, I., & Henry, D. Self-desensitization and meditation in the reduction of public speaking anxiety. *Journal of Consulting and Clinical Psychology*, 1979, *47*, 536–541. (63)

Kockott, G., Dittmar, F., & Nusselt, L. Systematic desensitization of erectile impotence: A controlled study. *Archives of Sexual Behavior*, 1975, *4*, 493–500. (64)

Kostka, M. P., & Galassi, J. P. Group systematic desensitization versus covert positive reinforcement in the reduction of test anxiety. *Journal of Counseling Psychology*, 1974, *21*, 464–468. (65)

Lamontagne, Y., Gagnon, M. A., Trudel, G., & Boisvert, J. M. Thought-stopping as a treatment for reducing cigarette smoking. *The International Journal of the Addictions*, 1978, *13*, 297–305. (66)

Lando, H. Self-pacing in eliminating chronic smoking:

Serendipity revisited? *Behavior Therapy*, 1976, *7*, 634–640. (67)

Lent, R. W., & Russell, R. K. Treatment of test anxiety by cue-controlled desensitization and study-skills training. *Journal of Counseling Psychology*, 1978, *25*, 217–224. (68)

Levenberg, S. B., & Wagner, M. K. Smoking cessation: Long-term irrelevance of mode of treatment. *Journal of Behavior Therapy and Experimental Psychiatry*, 1976, *7*, 93–95. (69)

Levitz, L. S., & Stunkard, A. J. A therapeutic coalition for obesity: Behavior modification and patient self-help. *American Journal of Psychiatry*, 1974, *131*, 423–427. (70)

Lick, J. Expectancy, false galvanic skin response feedback, and systematic desensitization in the modification of phobic behavior. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 557–567. (71)

Lick, J. R., & Heffler, D. Relaxation training and attention placebo in the treatment of severe insomnia. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 153–161. (72)

MacDonald, M. L., Lindquist, C. U., Kramer, J. A., McGrath, R. A., & Rhyne, L. D. Social skills training: Behavior rehearsal in groups and dating skills. *Journal of Counseling Psychology*, 1975, *22*, 224–230. (73)

Margolin, G., & Weiss, R. L. Comparative evaluation of therapeutic components associated with behavioral marital treatments. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 1476–1486. (74)

Marshall, W. L., Boutilier, J., & Minnes, P. The modification of phobic behavior by covert reinforcement. *Behavior Therapy*, 1974, *5*, 469–480. (75)

Marshall, W. L., Gauthier, J., Christie, M. M., Currie, D. W., & Gordon, A. Flooding therapy: Effectiveness, stimulus characteristics, and the value of brief in vivo exposure. *Behavior Research and Therapy*, 1977, *15*, 79–87. (76, 77, 78)

Marshall, W. L., Press, L., & Andrews, W. R. A self-administered program for public speaking anxiety. *Behavior Research and Therapy*, 1976, *14*, 33–39. (79)

Marshall, W. L., Stoian, M., & Andrews, W. R. Skills training and self-administered desensitization in the reduction of public speaking anxiety. *Behavior Research and Therapy*, 1977, *15*, 115–117. (80)

Marzillier, J. S., Lambert, C., & Kellett, J. A controlled evaluation of systematic desensitization and social skills training for socially inadequate psychiatric patients. *Behavior Research and Therapy*, 1976, *14*, 225–238. (81)

Mathews, A., Bancroft, J., Whitehead, A., Hackmann, A., Julier, D., Bancroft, J., Gath, D., & Shaw, P. The behavioral treatment of sexual inadequacy: A comparative study. *Behavior Research and Therapy*, 1976, *14*, 427–436. (82)

Mathews, A., & Rezin, V. Treatment of dental fears by imaginal flooding and rehearsal of coping behavior. *Behavior Research and Therapy*, 1977, *15*, 321–328. (83)

McConaghy, N. Aversive and positive conditioning treatments of homosexuality. *Behavior Research and Therapy*, 1975, *13*, 309–319. (84)

McCordick, S. M., Kaplan, R. M., Finn, M. E., & Smith, S. H. Cognitive behavior modification and modeling for test anxiety. *Journal of Consulting and Clinical Psychology*, 1979, *47*, 419–420. (85)

McGlynn, F. D., & McClaren, H. A. Components of desensitization in modification of fear among genuinely fearful subjects. *Psychological Reports*, 1975, *37*, 959–969. (86)

McReynolds, W. T., Lutz, R. N., Paulsen, B. K., & Kohrs, M. B. Weight loss resulting from two behavior modification procedures with nutritionists as therapists. *Behavior Therapy*, 1976, *7*, 283–291. (87)

Melnick, J., & Russell, R. W. Hypnosis versus systematic desensitization in the treatment of test anxiety. *Journal of Counseling Psychology*, 1976, *23*, 291–295. (88)

Mendonca, J. D., & Siess, T. F. Counseling for indecisiveness: Problem solving and anxiety management training. *Journal of Counseling Psychology*, 1976, *23*, 339–347. (89)

Miller, R. K., & Bornstein, P. H. Thirty-minute relaxation: A comparison of some methods. *Journal of Behavior Therapy & Experimental Psychiatry*, 1977, *8*, 291–294. (90)

Mitchell, K. R., Hall, R. F., & Piatkowska, O. E. A group program for the treatment of failing college students. *Behavior Therapy*, 1975, *6*, 324–336. (91)

Mitchell, K. R., & Orr, F. E. Note on treatment of heterosexual anxiety using short-term massed sensitization. *Psychological Reports*, 1974, *35*, 1093–1094. (92)

Moleski, R., & Tosi, D. J. Comparative psychotherapy: Rational-emotive therapy versus systematic desensitization in the treatment of stuttering. *Journal of Consulting and Clinical Psychology*, 1976, *44*, 309–311. (93)

Monti, P. M., Fink, E., Norman, W., & Curran, J. Effect of social skills training groups and social skills bibliotheraphy with psychiatric patients. *Journal of Consulting and Clinical Psychology*, 1979, *47*, 189–191. (94)

Moss, M. K., & Arend, R. A. Self-directed contact desensitization. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 730–738. (95)

Nemetz, G. H., Craig, K. D., & Reith, G. Treatment of female sexual dysfunction through symbolic modeling. *Journal of Consulting and Clinical Psychology* 1978, *46*, 62–73. (96)

Nietzel, M. T., Martorano, R. D., & Melnick, J. The effects of covert modelling with and without reply training on the development and generalization of assertive responses. *Behavior Therapy*, 1977, *8*, 183–192. (97)

Obler, M. Multivariate approaches to psychotherapy with sexual dysfunctions. *The Counseling Psychologist*, 1975, *5*, 55–60. (98)

O'Brien, G. T., & Borkovec, T. D. The role of relaxation in systematic desensitization: Revisiting an unresolved issue. *Journal of Behavior Therapy and Experimental Psychiatry*, 1977, *8*, 359–364. (99)

O'Brien, R. M. Negative practice and desensitization of anxiety about examinations. *Psychological Reports* 1976, *38*, 1147–1153. (100)

O'Brien, R. M., & Dickinson, A. M. Contingency factors in negative practice of smoking. *Psychological Reports*, 1977, *40*, 495–505. (101)

Odom, J. V., Nelson, R. O., & Wein, K. S. The differential effectiveness of five treatment procedures

three response systems in a snake phobia analog study. *Behavior Therapy*, 1978, *9*, 936–942. (102)

Ottens, A. J. The effect of transcendental meditation upon modifying the cigarette smoking habit. *Journal of School Health*, 1975, *45*, 577–583. (103)

Pederson, L. L., Scrimgeour, W. G., & Lefcoe, N. M. Comparison of hypnosis plus counseling, counseling alone, and hypnosis alone in a community service smoking withdrawal program. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 920. (104)

Pendleton, L. R., & Tasto, D. L. Effects of metronome-conditioned relaxation, metronome-induced relaxation, and progressive muscle relaxation in insomnia. *Behavior Research and Therapy*, 1976, *14*, 165–166. (105)

Rakos, R. F., & Schroeder, H. E. Fear reduction in helpgivers as a function of helping. *Journal of Counseling Psychology*, 1976, *23*, 428–435. (106)

Reinking, R. H., & Kohl, M. L. Effects of various forms of relaxation training on physiological and self-report measures of relaxation. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 595–600. (107)

Reister, B. W., Stockton, R. A., & Maultsby, M. C. Counseling the test anxious: An alternative. *Journal of College Student Personnel*, 1977, *18*, 506–510. (108)

Richards, C. S., & Perri, M. G. Do self-control treatments last? An evaluation of behavioral problem solving and faded counselor contact as treatment maintenance strategies. *Journal of Counseling Psychology*, 1978, *25*, 376–383. (109)

Richardson, F. C., & Suinn, R. M. Effects of two short-term desensitization methods in the treatment of test anxiety. *Journal of Counseling Psychology*, 1974, *21*, 457–458. (110)

Romano, J. L., & Cabianca, W. A. EMG biofeedback training versus systematic desensitization for test anxiety reduction. *Journal of Counseling Psychology*, 1978, *25*, 8–13. (111)

Rosen, G. M., Glasgow, R. E., & Barrera, M. A controlled study to assess the clinical efficacy of totally self-administered systematic desensitization. *Journal of Consulting and Clinical Psychology*, 1976, *44*, 208–217. (112)

Rozensky, R. H., & Bellack, A. S. Individual differences in self-reinforcement style and performance in self- and therapist-controlled weight reduction programs. *Behavior Research and Therapy*, 1976, *14*, 357–364. (113)

Russell, M. A. H., Armstrong, E., & Patel, U. A. Temporal contiguity in electric aversion therapy for cigarette smoking. *Behavior Research and Therapy*, 1976, *14*, 103–123. (114)

Russell, R. K., Miller, D. E., & June, L. N. A comparison between group systematic desensitization and cue-controlled relaxation in the treatment of test anxiety. *Behavior Therapy*, 1975, *6*, 172–177. (115)

Russell, R. K., & Wise, F. Treatment of speech anxiety by cue-controlled relaxation and desensitization with professional and paraprofessional counselors. *Journal of Counseling Psychology*, 1976, *23*, 583–586. (116)

Russell, R. K., Wise, F., & Stratoudakis, J. P. Treatment of test anxiety by cue-controlled relaxation and systematic desensitization. *Journal of Counseling Psychology*, 1976, *23*, 563–566. (117)

Saccone, A. J., & Israel, A. C. Effects of experimenter versus significant other-controlled reinforcement and choice of target behavior on weight loss. *Behavior Therapy*, 1978, *9*, 271–278. (118)

Shaw, B. F. Comparison of cognitive therapy and behavior therapy in the treatment of depression. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 543–551. (119)

Shaw, D. W., & Thoresen, C. E. Effects of modeling and desensitization in reducing dentist phobia. *Journal of Counseling Psychology*, 1974, *21*, 415–420. (120)

Sherman, A. R., Mulac, A., & McCann, M. J. Synergistic effect of self-relaxation and rehearsal feedback in the treatment of subjective and behavioral dimensions of speech anxiety. *Journal of Consulting and Clinical Psychology*, 1974, *42*, 819–827. (121)

Shoemaker, J. E., & Tasto, D. L. The effects of muscle relaxation on blood pressure of essential hypertensives. *Behavior Research and Therapy*, 1975, *13*, 29–43. (122)

Sipich, J. F., Russell, R. K., & Tobias, L. L. A comparison of covert densitization and "nonspecific" treatment in the modification of smoking behavior. *Journal of Behavior Therapy and Experimental Psychiatry*, 1974, *5*, 201–203. (123)

Sloane, R. B., Staples, F. R., Cristol, A. J., Yorkston, N. D., & Whipple, K. Short term analytically oriented psychotherapy versus behavior therapy. *American Journal of Psychiatry*, 1975, *132*, 373–377. (124)

Smith, J. C. Psychotherapeutic effects of transcendental meditation with control for expectation of relief and daily sitting. *Journal of Consulting and Clinical Psychology*, 1976, *44*, 630–637. (125)

Snowden, L. R. Personality tailored covert sensitization of heroin abuse. *Addictive Behaviors*, 1978, *3*, 43–49. (126)

Snyder, A. L., & Deffenbacher, J. L. Comparison of relaxation as self-control and systematic desensitization in the treatment of text anxiety. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 1202–1203. (127)

Spiegler, M. D., Cooley, E. J., Marshall, G. J., Prince, H. T., & Puckett, S. P. A self-control versus a counterconditioning paradigm for systematic desensitization: An experimental comparison. *Journal of Counseling Psychology*, 1976, *23*, 83–86. (128)

Thorpe, G. L. Desensitization, behavior rehearsal, self-instructional training and placebo effects on assertive-refusal behavior. *European Journal of Behavioral Analysis and Modification*, 1975, *1*, 30–44. (129)

Tiegerman, S., & Kassinove, H. Effects of assertive training and cognitive components of rational therapy on assertive behaviors and interpersonal anxiety. *Psychological Reports*, 1977, *40*, 535–542. (130)

Turner, R. M., & Ascher, L. M. Controlled comparison of progressive relaxation, stimulus control, and paradoxical intention therapies for insomnia. *Journal of Consulting and Clinical Psychology*, 1979, *47*, 500–508. (131)

Turner, S. M., & Adams, H. Effects of assertive training on three dimensions of assertiveness. *Behavior Research and Therapy*, 1977, *15*, 475–483. (132)

Vargas, J. M., & Adesso, V. J. A comparison of aversion therapies for nailbiting behavior. *Behavior Therapy*, 1976, *7*, 322–329. (133)

Weissberg, M. A comparison of direct and vicarious treatments of speech anxiety: Desensitization with coping imagery and cognitive modification. *Behavior Therapy*, 1977, *8*, 606–620. (134)

Weissberg, M., & Lamb, D. Comparative effects of cognitive modification, systematic desensitisation and speech preparation in the reduction of speech and general anxiety. *Communication Monographs*, 1977, *44*, 27–36. (135)

Wincze, J. P., & Caird, W. K. The effects of systematic desensitization and video desensitization in the treatment of essential sexual dysfunction in women. *Behavior Therapy*, 1976, *7*, 335–342. (136)

Wolfe, J., & Fodor, I. Modifying assertive behavior in women: A comparison of 3 approaches. *Behavior Therapy*, 1977, *8*, 567–574. (137)

Woolfork, R. L., Carr-Kaffashan, L., McNulty, T. F., & Lehrec, P. M. Meditation training as a treatment for insomnia. *Behavior Therapy*, 1976, *7*, 359–365. (138)

Wright, J. C. A comparison of systematic desensitization and social skill acquisition in the modification of a social fear. *Behavior Therapy*, 1976, *7*, 205–210. (139)

Wright, W., Morris, K. T., & Fettig, B. Comparative effects of social skill development. *Small Group Behavior*, 1974, *5*, 211–221. (140)

Wroblewski, P. F., Jacob, T., & Rehm, L. The contribution of relaxation to symbolic modeling in the modification of dental fears. *Behavior Research and Therapy*, 1977, *15*, 113–117. (141)

Zaichkowsky, L. D., & Kamen, R. Biofeedback and meditation: Effects on muscle tension and locus of control. *Perceptual and Motor Skills*, 1978, *46*, 955–958. (142)

Ziesat, H. A., Rosenthal, T. L., & White, G. M. Behavioral self-control in treating procrastination of studying. *Psychological Reports*, 1978, *42*, 56–69. (143)

Received August 17, 1981 ■