1988

Review of Cognitive-Behavioral Therapy with Troubled Children: A Meta-Analysis

Teresa Fuhrman
Loyola University Chicago

REVIEW OF COGNITIVE-BEHAVIORAL THERAPY WITH TROUBLED CHILDREN: A META-ANALYSIS

by

Teresa Suriano Fuhrman

A Thesis Submitted to the Faculty of the Graduate School of Loyola University of Chicago in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

October

1988

ACKNOWLEDGEMENTS

I wish to thank all those who have helped with the completion of this thesis. I am grateful to Joseph A. Durlak, Ph. D., thesis director, for his guidance throughout this long process. I would like to thank Fred B. Bryant, Ph. D. for his suggestions and direction concerning the nuances of meta-analytic technique.

Finally, I would like to thank my family for their encouragement, support, and patience, especially my husband, Stephen. His love and confidence in me are invaluable.

VITA

The author, Teresa Suriano Fuhrman, is the daughter of Francis Joseph Suriano, Ph. D. and June Mary (Hanson) Suriano. She was born November 19, 1962. Her elementary education was obtained in a Catholic school in Scottsdale, Arizona. Her secondary education was obtained in a Catholic high school in Phoenix, Arizona.

Ms. Fuhrman enrolled at the University of Arizona in September, 1981. In August, 1985, she graduated with the degree of Bachelor of Science in Psychology, with High Distinction. In 1984, while attending the University of Arizona, she was elected a member of Phi Beta Kappa.

In September, 1985, Ms. Fuhrman entered Loyola University of Chicago as a graduate student in the Clinical Division of the Department of Psychology.

TABLE OF CONTENTS

Page

LIST OF TABLES

CONTENTS OF APPENDICES

INTRODUCTION

Literature reviews have long served a significant purpose in scientific research. Reviews provide a format for summarizing the current state of the art in a field, indicating where gaps of knowledge exist, and establishing the "stubborn, dependable relationships that regularly occur" (Cook and Leviton, 1980, p. 449) despite variations across individual research projects.

Traditionally, literature reviews have used a narrative format. Typically a reviewer drew up a list of relevant studies, perhaps examined them for methodological quality, and then counted the number of relevant studies which confirmed or disconfirmed the relationships in question. The basis for conformation or disconfirmation was often reported statistical significance of measures. After all the studies had been divided on this basis, the side with the greatest number of studies won (Cook and Leviton, 1980).

Problems with Narrative Reviews

While the necessity of literature reviews is indisputable, the format for conducting the review is an issue of debate. The central problem with narrative reviews is the discrepancy between the amount of

information available and the amount which can be comprehended, organized, and integrated easily. In some areas of study, the amount of available research has multiplied beyond the level where it can be comprehended easily (Glass et al., 1981) when presented in the traditional narrative format. "The reviewer is even less able to absorb the sense of a hundred research studies than is an observer able to scan one hundred test scores and, without reliance on statistical methods, absorb the sense of their size and spread and correlations" (Glass, McGaw, & Smith, 1981, p. 14).

Another common complaint about narrative reviews is that important information is ignored. Light and Smith (1971) argue that the magnitide and direction of a relationship may be ignored in favor of a simplistic count of statistically significant findings in applicable studies. This simple frequency count strategy is problematic when trying to interpret studies with "mixed" results, i.e. when some outcome measures are positive and some are negative. Problems also arise when trying to interpret a group of studies with discrepant findings. The task of evaluating the effect of combinations of variables across studies is simply

too overwhelming to undertake using only frequency charts and short narrative descriptions of studies. As a result, reviewers often limit the focus of their analysis on a narrowed part of the available research, seldom clarifying how this limited sample was selected (Glass et al., 1981). Such a strategy gives rise to the criticism that conclusions in narrative reviews are subjective and not representative of the field.

Another problem with narrative reviews in fields where studies are prolific is the difficulty in assimilating conclusions of several reviews, each covering a different, or occasionally the same, aspect of a relationship. In other words, a "review of the reviews" may become necessary in order to understand what is known and where the gaps in knowledge are in understanding a relationship or area of research.

Meta-Analysis

Because of the need for integrative literature reviews involving all relevant research, meta-analysis has begun entering mainstream psychological journals. The essential characteristic of the meta-analytic review is the statistical analysis of summary findings (Bangert-Drowns, 1986). Meta-analytic technique offers investigators another way to synthesize and evaluate a

body of literature. By gathering together a large and cohesive body of research and converting the varied results to a common metric, comparisons across studies are feasible. The goal is to determine what the available research says about treatment effectiveness.

The reviewer will typically identify a group of studies that investigate the same research question. The diversity present in psychotherapy outcome studies is assumed to be due to differences in subject characteristics, experimenters, settings, and definition and focus of treatment. The meta-analyist then tries to answer the question, "Are some regular patterns discernable in a body of studies on a given topic that show divergent outcomes?" (Bangert-Drowns, 1986, p. 338). With outcome measures converted to a common metric, the reviewer can quantitatively examine variables which, based on either theory or logical argument, may mediate psychotherapy outcome.

Relative Advantages of Meta-Analysis

Recently, meta-analysis has become a popular method in psychotherapy research reviews. One explanation for the popularity of meta-analysis is that this procedure can be seen as responding to the limitations of a narrative review. The major advantage of meta-analysis

over narrative reviews is the use of quantitative method for ordering and integrating a collection of different treatment studies (Cook and Leviton, 1980; Glass et al., 1981). This advantage increases as the number of studies to be reviewed increases.

Because of its quantitative nature, meta-analysis offers the advantage of being able to more objectively assess treatment effectiveness and explore relationships among variables and interactions (Cook and Leviton, 1980). Therefore, meta-analysis addresses the complaint that narrative reviews ignore important information and are subjective in drawing conclusions. Because narrative reviews tend to focus on statistically significant results and ignore nonsignificant trends and mixed results, their conclusions may be more conservative than conclusions of a meta-analytic review (Cook and Leviton, 1980). Meta-analysis considers the direction and magnitiude of outcome measures in all studies; therefore, effect size of treatments which are nonsignificant in the predicted direction can more directly contribute to the review's overall findings. This is especially important when one considers that Cohen (1962) found that statistical power in most social psychological research is relatively low.

Meta-analysis also attempts to address the complaint that reviews are subjective in the selection of studies by explicitly stating the criteria for inclusion of a study and method of searching the literature for appropriate studies. Because meta-analysis is relatively new and has the potential for wide and varied application, criteria for inclusion and search method for conducting a meta-analysis are areas of debate among meta-analyists. However, while meta-analysts do not agree on the exact methodology in each instance, they typically agree on the need to be explicit about procedural decisions (Bullock and Svyantek, 1985).

Criticisms of Meta-Analysis

While the meta-analytic procedure for literature reviews is seen as responding to some of the limitations of traditional reviews, the procedure has been criticized on several grounds. Narrative reviews have been criticized for covering too few studies, while meta-analytic reviews have been criticized for trying to do too much and losing valuable information with overly broad constructs. The quality of studies included in a meta-analytic review, the extensiveness of the literature search, and the unit of analysis are all

areas of criticism and debate.

Mixing "Apples" and "Oranges"

The most common criticism of meta-analysis is that it is illogical because it attempts to mix unlike findings; that is, it mixes apples and oranges. However, perhaps this issue is raised because meta-analytic and narrative reviews tend to be different in definition and scope of their constructs. Cook and Leviton (1980) argue that there is a bias in favor of narrowly defined constructs with narrative literature reviews while meta-analytic reviews tend to use more broadly defined constructs. Glass (1978) argues it is best to analyze the data using the construct of "fruit" and let the data decide whether this construct should be further divided into "apples" and "oranges". To stretch this analogy even more, one could characterize the constructs in individual research studies as "California oranges ", "Florida oranges", "green apples", and "red apples"; the narrative review constructs as "oranges" or "apples"; and the meta-analytic review constructs as "fruit". A certain amount of clustering constructs together always takes place in psychotherapy outcome reviews and the precision of individual studies is lost while clarity

is gained in "the big picture". Therefore, the researcher must decide upon the purpose of the review and define constructs accordingly (Bangert-Drowns, 1986).

While meta-analytic reviews of psychotherapy outcome tend to use broad constructs, critical differences among studies can be overlooked with overly broad generalizations of data (Presby, 1978). For example, Glass and Smith (1977) found no significant differences between the behavioral and nonbehavioral psychotherapies. This meta-analysis grouped studies into two "superclasses". The behavioral superclass included Gestalt, systematic desensitization, implosion, and behavior modification. The nonbehavioral superclass included rational-emotive, transactional analysis, Adlerian, Freudian, psychoanalytic, and Rogerian therapies (Glass & Smith, 1977). The use of these very broad categories may have cancelled important differences between psychotherapies which led to the conclusion that there were minimal real differences between psychotherapies (Bangert-Drowns, 1986).

The Effects of Methodological Quality

Another controversy in meta-analysis concerns the quality of studies reviewed. Quality refers to factors

such as subject sampling strategy, sample size, use of a control group and random assignment, outcome measure choice and quality, and the particular levels of significance used (Bullock & Svyantek, 1985). Some researchers feel that the quality of the studies appropriate for a review should be evaluated first and that studies judged to be of poor quality should be excluded from the analysis. This follows the "garbage in-garbage out" philosophy, where conclusions drawn from research are only as valid as the evidence upon which they are based (Eysenck, 1978; Mansfield & Busse, 1977).

On the other hand, Glass and his colleagues (1981) have repeatedly argued that meta-analysts should not exclude studies on the basis of methodological quality because the "strategic combination" of studies with different weaknesses can cancel each other out and yield unbiased results. Glass (1978) states that when studies have been excluded on any grounds, it is possible that the sample has become biased.

Bryant and Wortman (1984) assert that exclusion decisions about study quality are guided by two concerns: the experimenter's priorities regarding internal and external validity, and the range of

experimental rigor in the area of interest. Further, they suggest different courses of action based on the range of experimental rigor present in the area of interest. When there is a wide range of study quality, for example, Bryant and Wortman (1984) recommend using all available studies in the meta-analysis and developing a way to code studies according to quality. Studies judged to have higher methodological quality can be compared with studies judged to be of poorer quality and the experimenter can evaluate the degree to which study outcome is related to methodological quality. When the range of study quality is restricted, such as when all available studies are non-randomized quasi-experiments, they recommend selecting only the studies judged to have the highest quality. With no baseline of high quality studies to use in a comparison, it is impossible to speculate about the effects of methodological quality on the therapeutic outcome. However, Fiske (1983) has criticized the idea of using methodological quality as a part of the study selection criteria, saying that there is a limited consensus on standards of assessing methodological quality. Because of this limited consensus, exclusion of studies based on this criteria creates an illusion

of agreement about what methodological quality is and biases the review findings.

Presently the most straightforward manner to proceed in light of the debate regarding methodological quality is to clearly and explicitly state in the review the criteria for inclusion and exclusion and provide a complete listing of all studies included and excluded. Informed readers can then decide for themselves the extent to which violations of internal and external validity are a concern. Further, if methodological quality is somehow defined, quantified, and coded, all studies included in a meta-analysis can be empirically examined to determine if differences in study quality as defined by the reviewer are related to differences in study outcome (Bangert-Drowns, 1986).

Another debate in meta-analysis focuses on the extensiveness of the literature review. Glass and colleagues (1981) feel that:

> locating studies is the stage at which the most serious form of bias enters a meta-analysis, since it is difficult to assess the impact of a potential bias. The best protection against this source of bias is a thorough description of the procedures used to locate the studies that were

> found so that the reader can make an intelligent assessment of the representativeness and completeness of the data base for a meta-analysis (p. 57).

Analyses suggest that the therapeutic effects of published studies are often higher than those found in dissertations (Bangert-Drowns, 1986; Smith, 1980). This phenomenon is called "publication bias". To control for publication bias, Light and Pillemer (1985) suggest an intensive search not only of published literature, but also of unpublished literature consisting of conference presentations and indices of research. They also suggest writing to researchers involved in the field of interest in the hopes of obtaining unpublished literature.

Unit of Statistical Analysis

A final methodological debate in meta-analysis relates to the unit of analysis. Glass and colleagues (1981) support assigning an effect size to every outcome measure of each study. Using this procedure, a study will contribute the number of effect sizes equal to the number of outcome measures. Consequently, certain studies with many outcome measures can have a strong influence on the meta-analysis, biasing the

results. Critics of using the outcome measure as the unit of analysis often prefer to use the study as the unit of analysis (Landman & Dawes, 1982; Mansfield & Busse, 1977; Wortman & Bryant, 1985). This is done by calculating an effect size for each outcome measure in a study and then averaging these measures to obtain a single effect size for each study. With each study contributing only one effect size, the problem of nonindependence of effect sizes is avoided allowing reviewers greater confidence in the use of statistical analysis. Because the number of effect sizes is equal to the number of studies, all studies are equally represented and weighted, allowing greater ease in interpretation.

In summary, literature reviews serve an important function in assisting researchers to assess the current state of the art and direct further inquiry. While the most prominent form of review is still narrative, recently the application of meta-analytic techniques have become more common. Meta-analytic technique increases the reviewer's arsenal of investigative tools and makes possible a more thorough, quantitative analysis in research review. The benefit of meta-analytic technique increases with an increase in

the amount of research in a given area of inquiry by providing a common metric for the unit of analysis and for quantitative evaluation of treatment effects.

The purpose of this thesis is to conduct a meta-analytic review of cognitive-behavioral therapy research with children. Before attempting a meta-analytic review of this research, it is necessary to discuss past reviews in this area and to define cognitive-behavioral therapy.

Cognitive-Behavioral Therapy

Cognitive-behavioral therapies (CBT) represent a synthesis of traditional and behavioral therapies (Kendall & Hollon, 1979). More traditional therapies focus on acquiring insight into problems but generally do not directly manipulate processes to produce behavioral change while behavioral therapies emphasize systematic interventions and observable phenomena and ignore unobservable processes. "Cognitive-behavior therapy accepts the processes (e.g., belief systems, expectancies, attributions) as basic data, and seeks also to fashion interventions and assess their effectiveness on the basis of sound scientific principles" (Kendall & Hollon, 1979, p. XV).

The distinguishing characteristic of

cognitive-behavioral therapy is an emphasis on thinking processes (Kendall, 1981) as opposed to an emphasis on teaching specific behaviors (behavior therapy) or uncovering internal conflicts and motivations (dynamic therapy). Cognitive-behavioral therapies highlight the importance of integrated and adaptive behavior while emphasizing cognitive change as the focus of treatment (Kendall, 1981). Therefore the focus is on the need to modify thinking processes by teaching the child effective adjustment strategies. While there are many different techniques subsumed in CBT, these techniques teach a child to use mediating responses that exemplify a general strategy for controlling behavior in a variety of situations (Gresham, 1981). Some believe that the implicit assumption of CBT is that treatment changes thoughts and problem-solving processes, which in turn alters behavior (Gresham, 1985). Cognitive-behavioral therapy includes interventions such as coaching, self-instruction training, attribution retraining, and problem-solving training. These interventions have been used with children in areas such as self control training, assertiveness training, social skills training, reduction of social isolation, reduction of hyperactive and aggressive behavior, and increasing

motivation (Cole & Kazdin, 1980; Gresham, 1981, 1985; Gresham & Lemanek, 1983; Hobbs, Moguin, Tyroler & Lahey, 1980; Kendall, 1981; Kendall & Hollon, 1979; Meador & Ollendick, 1984; Pearl, 1985; Urbaine & Kendall, 1980; Whalen, Henker, & Hinshaw, 1985; Wilson, 1984).

Cognitive-Behavioral Therapy Techniques

CBT is difficult to define, therefore it follows that there are many different techniques encompassed by CBT treatment. Some of these techniques include: self-instruction training, task or social oriented problem-solving training, role-playing, coaching, modeling, rewards, social skills training, attribution retraining or cognitive restructering, cognitive social learning, or social cognition training, and imagry. While CBT techniques share the assumption that the focus of treatment should be on what the child thinks in various situations, they differ in important ways.

Self-instruction training (Meichenbaum, 1977; Meichenbaum & Goodman, 1971) focuses on teaching children to guide their behavior by talking to themselves following a basic format. Many variations of self-instruction training exist based on the child's developmental level and the particular task or situation

the self-verbalizations apply, but the training involves learning to covertly ask: What is my problem; What is my plan; Am I using my plan; and How did I do? (Camp, Blom, Herbert, van Doorninck, 1977). Generally, treatment begins with the therapist moving through the task or situation and speaking out loud with the child watching (modeling), then the child moves through the task with the therapist helping (practicing). Next the therapist whispers the self-instructions with modeling and practicing repeated until the self-instructions are covert rather than overt. Coping with mistakes is also modeled throughout the therapy.

Problem-solving strategies were introduced into CBT by D'Zurilla and Goldfried (1971) and applied to tasks and activities. Essentially, problem-solving training follows the following steps: develop an orientation to recognize the problem; carefully define the problem and what needs to be done; generate alternative plans of action; chose a plan of action by evaluating all those generated in terms of gains and losses; and finally, evaluate the outcome the of plan of action (Cole & Kazdin, 1980). Spivak and Shure (1974) modified problem-solving strategies and developed a program specifically designed to improve children's peer

relationships. Spivak and Shure's (1974) social problem- solving technique involves four steps: teaching children to identify problems in interpersonal relationships; generate alternative solutions to the problems; anticipate the consequences of the various solutions; and make a step-by-step plan to carry out the chosen solution (Cole & Kazdin, 1980). Social cognition training, such as affective education and persepective training, can be a useful adjunct to social problem-solving training. Children are taught to broaden their understanding of their emotions in affective education, and to view a situation through more than one perspective in perspective training.

Several other techniques are frequently used as adjunct components in CBT. Role playing and coaching by therapist or peer provide the children with an opportunity to practice and improve their understanding of and use of skills taught in the CBT intervention. Elements of behavioral therapy such as shaping and reinforcing behavior with modeling, social praise, and concrete rewards are also sometimes added to a CBT intervention.

CBT interventions based on attribution retraining focus on changing overt and covert self-statements which

are believed to influence subject performance. Early work in attribution retraining was done by Dweck (1975) and involves helping children change their attributions for success and failure. For instance, an attribution retraining study may focus on changing a students attribution of sucess from luck to effort, and attributions of failure from stupidity to lack of sufficient effort. Attribution retraining also includes cognitive restructuring based on the work of Ellis (1963) which focuses on changing negative, anxiety producing self-statements to coping, self-reinforcing self-statements (Pearl, 1985).

Some literature reviews have included symbolic modeling (Gresham, 1985; Hobbs et al., 1980) as a form of CBT. Symbolic modeling is a form of treatment which involves showing children a film or videotape narrated in the first or third person. The film or videotape focuses on behavior which the researchers wish to change, such as showing socially withdrawn preschoolers a 23 minute film of children engaged in social behaviors (O'Connor, 1969).

The final technique frequently found in the CBT literature is social skills training. Social skills training involves teaching children who are presumed to

have a deficit in a certain skill. Behaviors such as giving and receiving positive social reinforcement, greeting others, nonaggressive responses to social frustration, and learning how to join in an activity are all considered social skills. For example, Gresham and Nagle (1980) taught cooperative and communication skills to socially isolated children through symbolic modeling and coaching.

Past Reviews of Research on Cognitive Behavioral Therapy

While there have been several reviews of cognitive-behavioral therapy with children, an integration of the findings of these reviews is difficult. Reviewers have often evaluated different studies, focusing on different aspects of CBT with differing samples. For instance, Gresham (1985) reviewed studies of social skills training for children which utilized modeling and coaching techniques; while Pearl (1985) reviewed of studies aimed at increasing academic motivation by changing attributions of academic success and failure. Not suprisingly, Gresham and Pearl reached different conclusions about what variaable may have impacted on treatment effectiveness.

Gresham (1985) concluded that some social skills training programs with cognitive-behavioral techniques,

particularly modeling and coaching, are effective in teaching social skills to children. However, he believed the efficacy of self-instruction training and social problem-solving techniques for social skills training had not been demonstrated. These latter studies tend to include less direct measures of socially relevant behavior, making conclusions about the impact of the interventions more difficult. Gresham's conclusion concurs with that of Hobbs and colleagues (1980) and of Lahey and Strauss (1982).

On the other hand, Pearl (1985) concluded that CBT in attribution retraining was clearly effective in changing attributions made by children about the causes of their academic successes and failures. Further, Pearl (1985) outlined specific variables which research had indicated effected treatment outcome such as the child's ability level, phrasing and timeing of attribution messages, and how clearly the message was tied to the students' success or failure. She also made suggestions about the direction of further research.

Urbain and Kendall (1980) reviewed the specific technique of social-cognitive problem-solving with children and, unlike Gresham (1985), found equivocal but

encouraging evidence for treatment effectiveness. While they felt the research was encouraging, they noted problems with the research such as a reliance upon nonclinical samples and the frequent absense of measures of overt behavioral adjustment. Urbaine and Kendall (1980) believed that the research indicated the social problem-solving therapy was effective, but stated that research was still needed to "distill the active ingredients of complex multi-faceted training programs that are responsible for reported treatment effects" (p. 138). In other words, they felt the cognitive components of the cognitive-behavioral interventions had not been proven essential to therapeutic change. Similarly, Presley (1979) also felt that relatively few statements could be made about the clinical utility of self-instructional training. He believed that little is known about the treatment variables contributing to the efficacy of the procedures due to the limited range of problems to which the technique had been applied.

Wilson (1984) reviewed self instrucion training as part of the literature on self-control treatment for aggressive children; while Gresham (1985) reviewed some conceptually similar research in his review of social skills training. Both authors had similar conclusions.

Wilson (1984) concluded that self-instruction training for self control had demonstrated effectiveness on tasks and tests, but did not demonstrate effectiveness in influencing classroom behavior. Similarly, Gresham (1985) did not feel that the cognitive elements social skills training had been demonstrated on socially relevant measures of behavior change.

In contrast to the conclusions drawn by Gresham (1985) and Wilson (1984), Duzinski (1987) conducted a meta-analysis of cognitive-behavioral strategies applied in an educational setting to remediate academic performance or modify classroom behavior, and found moderate treatment effectiveness. Based on 45 studies, Duzinski (1987) found an overall treatment effect size of 0.75, with a treatment effect size of 0.36 within the placebo control comparisons.

Duzinski (1987) found treatment more effective when it focused on remediating task related skills such as mathematics or reading than when treatment focused on behavioral adjustment such as reducing impulsivity or aggression. Special education students, with the exception of hyperactive/impulsive and aggressive children, benefitted more from CBT than regular education students and treatment effects were more

durable for students of average to below average ability. Duzinski (1987) found no clear differences in treatment effectiveness for studies which focused on the following treatment components: verbal self-instruction, problem-solving, modeling/coaching, overt or covert verbalizations, coping skills, or imagry. Further, he found a relationship between treatment effectiveness and subject age. When studies were categorized by subject age (three to six years, seven to ten years, and eleven and older), a trend existed such that studies with subjects in the middle age group tended to have smaller treatment effects.

While Duzunski's (1987) meta-analysis focused on cognitive-behavioral strategies in educational interventions Casey and Berman (1985) conducted a meta-analytic review of all psychotherapy with children, excluding academic interventions. Casey and Berman (1985) reviewed seventy-five psychotherapy studies with children, including fourteen studies of CBT with children. Over all therapy studies, they found a treatment effect size of 0.71, with behavioral, cognitive-behavioral, and non-behavioral studies having an estimated effect size of 0.91, 0.81, and 0.40 respectively (Casey & Berman, 1985). They found the

behavioral therapies produced significantly greater treatment effects than did nonbehavioral studies, they felt this could be explained by measurement variables, and they found the superiority of behavioral treatments was eliminated when outcome measures similar to treatment activities were removed from effect size calculations (Casey & Berman, 1985). No significant difference was found in outcomes of individual and group treatment, nor were therapist experience, education, or gender systematically related to treatment success.

Casey and Bermann (1985) were unable to address the issue of treatment effectiveness with different subject problems, because the behavioral and nonbehavioral therapy studies tended to examine different target problems. However, Casey and Berman (1985) found the treatment effectiveness for different problems varied, with treatment for problems in social adjustment less effective than treatments for other target problems Further, Casey and Berman (1985) looked at subject characteristics such as percentage of boys (versus girls), age, intellectual functioning, and school grade in relation to treatment efficacy and found only percent of boys in the sample correlated with psychotherapy outcome. Studies with a higher proportion

of boys versus those with a smalleer proportion of boys had smaller treatment effects (Casey & Berman, 1985).

Although many reviews have been conducted, drawing conclusions about the efficacy of cognitive-behavioral interventions for children with problematic behavior is difficult because the present reviews tend to focus on different types of cognitive-behavioral therapy, different treatment goals, and different subject samples. This makes it equally difficult to assess what subject and treatment variables moderate the effectiveness of various interventions and what interactions between these variables may exist. A meta-analytic review of cognitive-behavioral therapy with children may clarify these currently foggy issues by more efficiently and effectively summarizing the current literature.

Hypotheses

The purpose of this thesis is to review cognitive-behavioral therapy for children using meta-analytic techniques. The following questions will be investigated:

1.) What is the overall effectiveness of cognitive-behavioral therapy with children?

a.) Treatment is predicted to have a positive

effect consistent with effects found for adult psychotherapy outcome research and consistent with the work of Casey and Berman (1985) and Duzinski (1987).

b.) Placebo control group versus treatment comparisons are expected to have smaller effect size estimates than no treatment control group versus treatment comparisons because placebo groups control for nonspecific treatment effects (Dush, Hurt, & Schroeder, 1983).

c.) Variables such as year of completion and publication status will be evaluated in relation to treatment effect, although the direction these variables may impact on treatment effect size cannot be predicted based on previous research.

2.) Are some forms of cognitive-behavioral therapy with children more effective than others?

a.) Present literature does not clearly indicate that one form of CBT is more effective than others, therefore this is a preliminary investigation. However, several reviewers are cautious regarding the efficacy of social problem-solving and problem-solving interventions (Gresham, 1985; Urbaine & Kendall, 1980; Wilson, 1984).

b). Conservative clinical lore suggests that

longer treatments will be more effective than shorter treatments, however this has not been substantiated by recent mata-analyses (Casey & Berman, 1985; Duzinski, 1987; Dush et al, 1983). The relationship between length of treatment and treatment outcome will be explored, with longer treatments predicted to yield greater treatment effects.

c.) Treatment is not expected to vary based on treatment setting or method of delivery.

3.) Is cognitive-behavioral therapy more effective with some children than others?

a.) Children who externalize their symptoms are expected to show greater improvement than children who internalize their symptoms, because it is assumed that, with treatment, externalizers will be mastering an adaptive skill they were not proficient with prior to treatment (Meichenbaum, 1977; Meichenbaum & Goodman, 1971).

b.) The relationship between subject age and treatment effectiveness is difficult to predict. While older children with greater cognitive skills are expected to benefit most from the verbal focus of CBT, Duzinski (1987) found a nonlinear relationship between age and treatment effect.

c.) Consistent with Casey and Berman's (1985) findings, studies with a higher proportion of boys are predicted to have smaller treatment effects.

d.) Length of treatment and severity of subject's symptoms are expected to interact. Subjects with clinically significant problems are expected to show greater improvement than subjects without clinically significant problems when treatment is long.

4.) Characteristics of outcome measures are expected to effect treatment outcome.

a.) Measures of cognitive change and measures of behavioral change are expected to correlate within a study.

b.) Effect size estimates of zero are expected to be significantly smaller than effect size estimates based on means and standard deviations.

c.) Effect size estimates from study statistics other than means and standard deviations are not expected to be significantly different from effect size estimates based on means and standard deviations.

d.) In evaluating the type of outcome measure, behavioral observations and therapist ratings are predicted to yield the greatest estimates of effect size.

e.) In evaluating the source of data, therapist rating are expected to yield the greatest treatment effects. 5.) The quality of research is potentially related to treatment outcome, however the direction is not predicted.

METHOD

Literature Search Procedure

Three methods were used to locate potentially relevant studies: manual search of journals, review of reference lists from narrative reviews, and computerized search of Dissertation Abstracts International. First, the contents of fourteen psychological journals judged most likely to contain child psychotherapy research were manually searched, including volumes published from 1970 to March, 1987. The year 1970 was chosen as a starting point for the literature search because the several major early works in this area were published in the early 1970's (Dweck, 1975; Meichenbaum & Goodman, 1971; Spivack & Shure, 1974). The following journals were manually searched: Journal of Clinical Child Psychology, Journal of Abnormal Child Psychology, Journal of School Psychology, Journal of Consulting and Clinical Psychology, Journal of Applied Behavior Analysis, Behavior Therapy, Behavior Modification, Cognitive Therapy and Research, American Journal of Community Psychology, Elementary School Guidance and Counseling, Psychology in the Schools, School Counselor, Journal of Consulting Psychology, and Journal of Community

Psychology.

Reference lists from narrative literature reviews of CBT (e.g., Abikoff, 1985; Meador & Ollendick, 1984; Pearl, 1985; Urbaine & Kendall, 1980; Whalen, Henker, & Henshaw, 1985; Wilson, 1984) were also inspected; and a significant number of studies from these reviews had already been uncovered in the manual search of journals. Finally, dissertations were obtained through a computer search of Dissertation Abstracts International. This computer search resulted in nineteen dissertations which appeared to meet inclusionary criteria; nine dissertations were available and ten were unavailable. Ten additional dissertations were obtained through a computer search of Dissertation Abstracts International for a related project. While it is ideal to obtain unpublished studies from a variety of sources to examine the possibility of publication bias, time constraints did not permit this kind of extensive search. Therefore only unpublished dissertations were included in this review as a sample of unpublished studies.

Criteria for Review

Cognitive-behavioral therapy is defined as treatment which seeks overt behavioral change by teaching children to change thoughts and thought

processes in an overt, active manner (Kendall, 1981). This definiton of CBT excludes studies where the sole therapeutic component was symbolic modeling, because symbolic modeling is considered a covert rather than overt process. Restricting the definition of CBT to include only those interventions overtly seeking to change cognitive processes should result in a more theoretically homogeneous sample of CBT interventions.

The minimum requirements for inclusion of a study in this review are the following:

1.) The therapy reviewed must correspond to the above definition of cognitive-behavioral therapy.

2.) Cognitive-behavior therapy must be compared with a control group drawn from the same population as the treated subjects.

3.) The therapeutic intervention must be implemented with children who show a behavioral problem or are judged to be functioning poorly in comparison to their peer group. The problem does not have to be defined as a clinical syndrome, but must be judged problematic for the child. Studies aimed at purely academic outcomes and studies designed to enhance the performance of normally functioning children will be excluded.

4.) Studies must involve cognitive-behavior therapy

interventions where the child is the sole and direct client. Other forms of treatment such as family therapy, peer counseling, teacher consultation, and parent training are beyond the scope of this meta-analysis.

5.) Studies must use subjects with a mean age of thirteen years or younger.

6.) Studies must be conducted in North America and printed in the English language.

7.) Studies must be completed in the time period from 1970 to March, 1987.

In summary, 49 studies (32 published studies and 17 unpublished dissertations) were included in this meta-analysis. All studies which met the inclusionary criteria, and which were therefore included in the meta-analysis, are listed in Appendix A. Fourteen studies which met the inclusionary criteria, but which were excluded because the data were not amenable to meta-analytic techniques (i.e., no means and standard deviations could be obtained from them), are listed in Appendix B.

Coding Procedure

Ten studies were initially coded so that problems with the coding sheet could be identified and corrected.

All studies were then coded with the revised coding sheet (see Appendix C). Each study was coded with respect to 49 subject, treatment, and design feature characteristics and seven outcome measure characteristics. Most classifications are straightforward and require no explanation.

Several writers have recommended that the quality of research design should be evaluated to assess the relationship between study quality and outcome (Bryant & Wortman, 1984; Bullock & Svyantek, 1985; Light & Pillemer, 1984). In this review, all studies were rated according to: (1) the sample size in treatment groups, (2) the use of random assignment (or a check on pretreatment equivalence of groups), (3) attrition rates, (4) the presence of at least one normed or blinded behavioral outcome measure (5) the type of control group used, and, (6) reporting post-test data for all the instruments which are used in the study at pre-test. These items were examined in relation to treatment effect size and were also pooled to create a six point index of design quality. For the index of design quality, one point was awarded for each design feature met. The criteria are explained in the coding sheet found in Appendix C.

The target symptom for the intervention was coded based on clinical significance. All studies were coded on the following question: Is it reasonable to consider the subject's problem a clinically significant problem? The following options as answers to the above question were available: (a) yes, (b) no, and, (c) uncertain or unclear. Subject problems which would clearly warrant treatment at a local mental health agency or by a mental health worker were classified as clearly clinically significant. For example, the following studies were rated as containing a subjects sample which exhibited clearly clinically significant problems: subjects referred to a guidance center because of aggressive behavior problems (Cannavo-Antognini, 1979), subjects with a clinical diagnosis of Attention Deficit Disorder with Hyperactivity (Brown, Wynne, & Medenis, 1985), subjects with a specific learning disability (Thomas & Pashley, 1982), and subjects scoring in the top 10% on a standardized behavior rating scale of aggression and considered aggressive by adults working with the children (Hunter, 1985).

Target problems rated as not clinically significant include problems such as: subjects receiving a low score on a peer rating of desirability to work with or play

with the subject (La Greca & Santogrossi, 1980), and subjects scoring poorly on Kagan's (1966) Matching Familiar Figures Test.

When the degree of the subjects' symptomotology was not clearly and unequivocally of clinical significance, yet the subjects displayed some symptomotology, studies were classified as uncertain/unclear on the dimension of clinical significance of the target problem. An example of symptomotology rated as uncertain/unclear on the dimension of clinical significance is subjects labeled non-self-controlled by a teacher in response to the researcher request (Kendall & Braswell, 1982). It is not clear that the teachers would have considered the behavior exhibited by the subjects as problematic regardless of the researchers request that the teacher list children with non-self-controlled behavior.

Measure of Treatment Efficacy

Comparisons between treatment groups are expressed in terms of the standardized effect size (Cohen, 1977):

$$d = \frac{Mt - Mc}{Sd\ pooled}$$

where Mt is the mean of the treated group, Mc is the mean of the comparison group, and Sd pooled is the

pooled within group standard deviation. When this information was not available, an effect size was estimated using other reported statistics following procedures described by Holmes (1986) and Wolfe (1986). When the results for an outcome measure were not reported or were reported only as nonsignificant, the effect size for the outcome measure was conservatively estimated as zero. When data for a measure was incompletely reported in terms of subscale scores, subscale data which was provided was transformed to an effect size estimate, missing subscale scores were estimated as zero, and all effect size estimates for the subscales were averaged to produce one effect size estimation for the measure.

The zero estimates of effect size may dilute the average effect size and may reduce the variance in the sample of effect size estimates. In fact, Hedges (1982) found the sampling distribution for effect size slightly biased and he developed a weighted estimate of effect size to adjust for this problem. The mean effect size was adjusted using Hedges' (1982) formula for the mean unbiased effect size (du) and Rosenthal and Rubin's (1982) formula to estimate w:

$$du = \frac{\text{sum of } (wd)}{\text{sum of } w}$$

$$\text{where } w = \frac{2N}{8 + d(\text{squared})}$$

N is the total sample size in each study and d is the mean effect size for that study.

Two distinct meta-analyses were conducted. The first utilized a single effect size estimate for each study, obtained by averaging the effect size across all outcome measures and CBT treatment groups. Design features, and subject and treatment characteristics were then evaluated in relation to treatment effectiveness. The second meta-analysis explored outcome measure characteristics by calculating separate effect sizes for each outcome measure in each study.

Some studies used several treatment and/or control groups. For example, a study could compare cognitive-behavioral therapy and behavior therapy using both a placebo control group and a no treatment control group. In this case, several comparisons applicable to

this review are possible: cognitive-behavioral therapy versus behavior therapy; cognitive-behavioral therapy versus the placebo control group; cognitive-behavioral therapy versus the no treatment control group. Since each study contributed only one effect size, whenever a choice existed, the comparison between cognitive-behavioral therapy and the placebo control group was utilized. This comparison was expected to control for placebo-related treatment gains while providing the most straightforward assessment of the effects of the cognitive- behavioral therapy. When a study utilized more than one type of CBT, each CBT treatment group was compared to the control group.

Rater Reliability

Five studies were randomly drawn from the sample and independently coded by the author and the thesis chairman. One estimate of interrater reliability was obtained for each meta-analysis. Interrater reliability for the 49 variables per study pertaining to design, subject, and treatment characteristics was 80%. Interrater reilibility for the seven variables per measure focused on outcome measure charactristics was 89%.

RESULTS

Initial Analyses

Fourteen of the 49 studies provided information about more than one type of cognitive-behavioral treatment. Since there was no a priori basis upon which to select one form of treatment over another within a study for inclusion in the meta-analysis, each form of treatment was compared with the control group. The mean effect sizes for the comparisons were then analyzed in two ways: (1) using only one effect size per study by averaging across the comparisons when a study contributed more than one treatment to control group comparison, resulting in 49 effect size estimates; and, (2) allowing each treatment versus control comparison to contribute one effect size, resulting in 63 effect size estimates.

Estimates of Average Effect Size

Table 1 displays the various estimates of effect size and the 95% confidence interval associated with each effect size. Using method 1, the mean effect size was 0.69 (*sd*= 0.94). The unbiased estimate of effect size (Hedges, 1982) using method 1 is 0.53. Removing one outlier effect size and recalculating with method 1

Table 1

Effect Size for the Sample

Method	Number of Comparisons	Mean (SD)	95% Confidence Interval	Fail-Safe N(a)
1	49	0.69 (0.94)	0.64 - 0.74	120 studies
1	48(b)	0.56 (0.42)	0.54 - 0.58	86 studies
1 Hedges adjusted d	49	0.53		81 studies
2	63	0.73 (1.13)	0.67- 0.78	167 studies
2 Hedges adjusted d	63	0.52		78 studies

Note.

(a) = Number of studies finding no significant effect that would be needed to reduce mean ES to nonsignificance.

(b) = with outlier removed.

results in an average effect size of 0.56 (sd= 0.42) across 48 studies. Based on method 2, the mean effect size was 0.73 (sd= 1.13) and an unbiased estimate of effect size of 0.52.

An effect size of 0.69 indicates that an person in the treatment condition performed 0.69 standard deviations above the average person in the control condition, or at the 75th percentile of the control condition. Thus, the average person receiving treatment showed more improvement than 75% of the people in the control group. An effect size of 0.73 indicates that the average person in the treatment condition performed at the 77th percentile of the treatment group.

Since both methods of effect size calculation yielded similar effect sizes, method 1 was used for all subsequent analyses. This is the more conservative approach because it controls for nonindependence of the comparisons. When the distribution of effect sizes was examined, one study (Kahl, 1985) yielded an effect size substantially different and higher than the remaining studies (ES= 6.51). Inspection of this outlier did not suggest anything unique about the study's design, subjects, or treatment. Nevertheless, a decision was made to conduct all subsequent analyses without this

outlier. Analyses including the outlier effect size are reported only when the presence of the outlier effect size significantly changes the findings.

Fail-Safe N

Regardless of how thorough a literature review is, it is unlikely that every applicable study will be uncovered. Rosenthal (1979) called this the "file drawer problem" or publication bias, because studies with nonsignificant findings are more likely than studies with significant findings to be left unpublished in a file drawer. Orwin's (1983) Fail-safe N estimates the number of studies with nonsignificant findings that would be needed to reduce an obtained finding to nonsignificance. A Fail-safe N was calculated for each estimate of effect size and is found in Table 1. Cohen's (1977) suggestion that an effect size of 0.20 is a small or nonsignificant effect size in the social sciences was utilized in these computations. Accordingly, 120 studies with nonsignificant findings would be needed to reduce the finding of a treatment effect size of 0.69 to nonsignificance and 81 studies would be needed to reduce the mean effect size of 0.53 to nonsignificance.

Preliminary Examination of Variance

The mean effect size was then tested for homogeneity of variance. Based on Rosenthal and Rubin's (1982) formula, the 2 $\underline{X}(48,1)=5496.06$, $\underline{p}< 0.001$. This finding indicates that the variance in the effect size is heterogeneous, and that study characteristics could be influencing effect size.

Preliminary analyses assessed differences between the effect size of published studies and unpublished dissertations, variations in effect size based on year of publication or completion, and type of control group to see if these study variables created significant differences in effect sizes. Throughout this report, results from $\underline{t}$-test analyses will be reported as using either a pooled variance estimate or a separate variance estimate when appropriate (i.e., when the variance for the groups are significantly different). For purposes of clarity in reporting, any $\underline{t}$-test utilizing a separate estimate of variance will be marked with an asterisk.

There was no significant difference in the effect size of published studies and dissertations, $\underline{t}(17.24*)=0.00$, $\underline{p}= 1.00$, two-tailed. Thirty-two published studies had a mean effect size of 0.56 ($\underline{sd}=$ 0.35), while 16 dissertations had a mean effect size of

0.56 (*sd*= 0.56.) Including the outlier effect size significantly changed the variance for the dissertations, (*M*= 0.96, *sd*= 1.63) although the comparison with published studies remained nonsignificant, *t*(14.57*)= -0.93, *p*= 0.37, two-tailed. Studies ranged in completion or publication date from 1971 to 1988, and this date was not significantly correlated with effect size, *r*(48)= -0.13, *p*= 0.38. As a result, further analyses were pooled across published studies and dissertations, and year of publication or completion because of the nonsignificant findings.

Because the placebo control comparisons were expected to have a smaller effect size than the no-treatment control comparisons, a one-tailed *t*-test was used to compare the effect size of no-treatment control comparisons to the effect size of placebo-control comparisons, *t*(46)= 1.16, *p*= 0.25. The placebo control versus therapy comparisons (*n*=34) resulted in a mean effect size of 0.52 (*sd*= 0.37) while the no treatment control group versus therapy comparison (*n*=14) resulted in a mean effect size of 0.67 (*sd*= 0.51.) Further analyses were pooled across type of control group because of this nonsignificant finding.

In summary, initial analyses revealed a moderate

treatment effect for CBT with significant variance in the sample of effect size estimates. Preliminary analyses indicate that publication status, date of completion, and type of control group used were not significantly related to treatment effect size.

Design Characteristics

Table 2 summarizes the effect size information on variables related to several design characteristics such as the general design of the study, the group assignment procedure, the consistency with which results of outcome measures were reported, and the use of standardized measures. Study designs included pretest posttest designs with a nonequivalent control group (n= 5), posttest only with a nonequivalent control group (n= 2), randomized true experiment (n= 35), or "other", e.g., matching (n= 6). The number of comparisons utilizing true experimental design in this sample was commendable. Type of design was not significantly related to effect size, $F(3,46)= 0.61$, $p= 0.61$.

Group assignment procedures included random assignment (n= 26), matching (n= 11), available intact groups (n= 9), voluntary self-selection (n=0), and other (n= 2). There was no significant difference in effect size based on the procedure used to assign

Table 2

Mean Effect Size based On Design Characteristics

	n	Effect Size (sd)
I. Type of Design		
pretest posttest w/ nonequivalent control group (NECG)	5	0.38 (0.18)
posttest only w/ NECG	2	0.79 (0.66)
randomized true experiment	35	0.60 (0.43)
other (e.g., matching)	6	0.50 (0.43)
II. Group Assignment Procedure		
random	26	0.52 (0.38)
matching	11	0.62 (0.42)
available intact	9	0.68 (0.53)
voluntary self selection	0	-- --
other	2	-- --
III. Did the study use random assignment or report pretreatment equivalence of groups?		
No	8	0.40 (0.23)
Yes	40	0.60 (0.44)

(continued)

Table 2 (continued)

	n	Effect Size (sd)
IV. Were all the same measures reported at pretest and posttest?		
No	8	0.39 (0.23)
Yes	40	0.60 (0.44)

subjects to treatment groups, $F(2,46)= 0.56$, $p= 0.58$. Forty of the 48 comparisons either utilized random group assignment or confirmed pretreatment equivalence, however, these studies were not significantly different from studies which did not report efforts to insure pretreatment group equivalence, $t(46)= -1.25$, $p= 0.22$, two-tailed.

The number of days between the end of treatment and the collection of posttest data was analysed in relation to effect size. Eighteen studies directly reported this information; twenty studies provided adequate information to estimate this length of time, and ten studies were not included in this analysis because the above information could not be ascertained or estimated. Suprisingly, the range of days between treatment and post-treatment measurement ranged from 0 to 30 days. Because many researchers would consider measurements occuring two weeks or greater following treatment as a follow-up measurement, the comparisons were split into two groups: (1) less than 2 weeks; and (2) greater than two weeks. These two categories were not significantly differently in effect size, $t(36)= 0.31$, $p= 0.38$, one-tailed.

Some researchers did not report data for every

outcome measure used in their study (n= 8). Whenever this happened, the effect size for that measure was conservatively estimated as zero, reducing the mean effect size estimate and variance. To evaluate how this effected the overall effect size estimate, studies which did or did not report data for all the instruments at pre- and posttesting were compared. A one-tailed t-test resulted in an nonsignificant trend, t(46)= -1.31, p= 0.10. This comparison was statistically significant when the outlier effect size was included, t(29.92*)= -1.71, p= 0.04; studies which did not report data for all instruments produced significantly smaller effect size estimates.

Possible ratings on the index of design quality ranged from zero to six, with higher scores indicating a study met more criteria judged as good study design characteristics. The criteria which composed the index of design quality were as follows: sample size, random assignment or pretreatment equivalence, attrition rate, at least one standardized or blinded outcome measure, type of control group, and use of the same instruments at pretest and posttest. Ratings of design quality ranged from 3 to 6, with a mean rating of 4.67 (sd= 0.98). Over half of the studies received a rating of 5

or 6, indicating they met all or all but one of the criteria for good study design. Because the ratings on index of design quality ratings were restricted, this variable was not analyzed in relation to effect size. Each criteria used to create this index, most of which have already been discussed, were analyzed in relation to effect size. None of the analyses were significant.

In summary, group assignment procedures, study design, number of days between the end of treatment and the collection of the posttest data, and the index of design quality were not significantly related to effect size. A nonsignificant trend was found such that studies which did not report results for the all measures they used yielded a smaller mean effect size than studies which reported results for all the measures they used.

Subject Characteristics

Subject characteristics such as age, gender, ethnicity, type of problem, source of subjects, and severity of the subject's problem were analysed in relation to effect size. These analyses are summarized in Table 3. It was impossible to analyze the results as a function of ethnicity of subjects because this information was reported in only sixteen of the 48

Table 3

Mean Effect Size based on Subject Characteristics

	n	Effect Size (sd)
I. Source of subjects recruited for the study		
inpatient/residential	8	0.70 (0.46)
outpatient	1	1.58
volunteers for special project	5	0.40 (0.40)
school	34	0.53 (0.39)
inpatient/residential and outpatient	9	0.80 (0.52)
II. Is it reasonable to consider the subject's problem a clinically significant problem?		
No	18	0.50 (0.42)
Yes	15	0.64 (0.50)
Uncertain	15	0.57 (0.32)

(continued)

Table 3 (continued)

	n	Effect size (sd)
III. Target problem		
social isolation	5	0.50 (0.39)
fear/phobia	3	0.58 (0.40)
anxiety	1	1.49
depression	1	0.55
other internalizing symptoms	1	0.58
hyperactivity/ impulsivity	15	0.57 (0.50
behavior management problem/aggression	5	0.77 (0.36)
psychotic/autistic	0	---- ----
other externalizing symptom	6	0.51 (0.35)
social skills, undefined	0	---- ----
mix of above or other	11	0.44 (0.36)
IV. Broadband syndromes		
overcontrolled	11	0.62 (0.42)
undercontrolled	26	0.59 (0.44)

studies. Of these sixteen, nine studies used subject samples in which the majority of the subjects were caucasian, five studies used subject samples in which the majority of the subjects were minority, and two reported using a sample with a balance of majority and minority subjects.

Subjects ranged in age from 5.5 to 13.0 years, with an average age of 9.5 years (sd= 1.65). The relationship between subject age and treatment effect size was examined in several ways. First, there was no significant linear correlation between the subject's age and treatment effectiveness, r(48)= 0.14, p= 0.33. Next, the studies were divided into three groups based on subjects' age: early school-aged children from 5 through 8 years (n=13), children from 9 to 11 years (n= 28), junior high school aged children ages 11 and older (n= 6). A nonlinear significant relationship was evident, F(2,45)= 4.89, p= 0.01. The mean effect size for studies based on subject age are as follows: 5 through 8 years (ES= 0.61), 9 to 11 years (ES= 0.46), and 11 years and older (ES= 0.99). Student-Newman-Keuls procedure at the 0.05 level revealed the studies with the junior high school aged children had a mean effect size significantly larger than the studies with children

ages nine to eleven years, at the 0.05 level. Duzinski (1987) found a similar nonlinear relationship between subject age and effectiveness of cognitive-behavioral strategies.

While studies had been divided into age groups based on general level of development, the above finding might have been accounted for by the fact that 28 of the 48 studies used subjects from nine to eleven years old. The relationship between subject age and effect size was explored again, this time with approximately equal numbers of studies in each age group: ages 5 to 9 years, 9.3 to 10 years, and 10.3 to 13 years. When studies were equally distributed across groups, a near significant relationship was found between subject age and effect size, $F(2,45) = 2.98$, $p = 0.06$. Student-Newman-Keuls procedure at the 0.05 level indicated the effect sizes for the older children were significanlty larger than for the children in the middle age group. While these age groups are hard to justify based on developmental logic, the relationship remained significant.

Casey and Berman (1985) found that studies which had a greater percent of boys were correlated with smaller overall treatment effectiveness. In the current

analyses, the subjects samples tend to be biased in the direction of using more boys than girls; the average study had a subject sample that was 72% male, with 12 of the 48 studies using only male subjects. There was no significant correlation between percent of boys versus girls in a study and treatment effect, $r(35) = 0.06$, $p = 0.74$, however, this variable was restricted in range.

Subjects were recruited for treatment from four different settings: inpatient or residential, outpatient clinic, newspaper advertising or other sources for volunteers, and schools. Treatment effectiveness differed significantly based on the source of subjects, $F(3,47) = 3.00$, $p = 0.04$. Inpatient, residential, and outpatient clinic subjects were then combined to form a larger cell for comparison with school children, resulting in a nonsignificant trend such that studies using clinic children had marginally greater treatment effect than studies using school children, $T(41) = 1.74$, $p = 0.09$. (See Table 3).

Did treatment effectiveness vary with the severity of the subjects' problem? The subjects in each study were rated on a 3- point scale of clinical significance of symptoms: "Is it reasonable to consider the subject's

problem a clinically significant problem?": (a) no (n=18); (b) yes (n=15), and; (c) uncertain or unclear (n=15). The studies which contained subjects with or without clearly clinically significant problems were compared and were not significantly different in effect size, T(31)= -0.89, p= 0.38, two-tailed. The primary problem exibited by the subjects was not significantly related to effect size, F(8,47)= 0.92 p= 0.51. (See Table 3 for the target problems and associated effect sizes). Because some of the original categories of target problems had very few studies, the data were collapsed into two categories based on Achenbach's (1978) concept of broadband syndromes of disorder. Subjects whose primary problems were defined as social isolation, fear, phobia, anxiety, depression, or another internalizing symptoms were clustered together as overcontrolling syndromes (n= 11). Subjects whose primary problems were defined as hyperactivity, impulsivity, behavior management problem, aggression, or other externalizing symptoms were clustered as undercontrolling syndromes (n= 26). Studies which treated children, some of whom could fit into overcontrolling syndromes and some of whom could fit into undercontrolling syndromes, were excluded (n=11).

It was hypothesized that undercontrolled subjects would show greater improvement than overcontrolled children, but the data do not support this hypothesis, $t(18) = -0.28$, $p = 0.39$, one-tailed. Table 3 contains the mean effect size of treatment based on overcontrolled and undercontrolled problems.

In summary, a nonlinear relationship was found between subject age and effect size, such that children aged 11 and older displayed significantly more treatment gains than children eight to eleven years old. The proportion of boys versus girls in a study, the setting from which the subjects were recruited for treatment, severity of subjects' problem, and nature of the subjects' problem were all unrelated to treatment effect size.

Treatment Characteristics

The treatment characteristics analyzed in relation to effect size include individual versus group therapy, length of treatment, treatment setting, and the type of CBT administered assessed in terms of treatment components. See Tables 4 and 5 for a summary of treatment characteristics in relation to effect size.

Length of treatment was examined in relation to treatment effectiveness. Treatments tended to be biased

Table 4

Mean Effect Size based on Treatment Characteristics

	n	Effect Size (sd)
I. Treatment modality		
individual	24	0.60 (0.39)
group	21	0.50 (0.40)
individual and group	3	0.76 (0.77)
II. Treatment Setting		
school	30	0.53 (0.39)
mental health clinic	2	1.22 (0.50)
residential	12	0.56 (0.44)
camp	1	0.64
combination of above or other	3	0.20 (0.14)

Table 5

Mean Effect Size based on Treatment Component

Component	n	Effect Size (sd)	T-test value	p
Task oriented problem-solving				
present	16	0.51 (0.40)	0.65	0.52
absent	32	0.59 (0.43)		
Social problem-solving				
present	14	0.59 (0.36)	-0.26	0.80
absent	34	0.55 (0.44)		
Self-instructions				
present	26	0.47 (0.34)	1.61	0.11
absent	22	0.67 (0.48)		
Role play				
present	22	0.54 (0.39)	0.43	0.67
absent	26	0.59 (0.44)		
Rewards				
present	10	0.50 (0.23)	0.53	0.60
absent	38	0.58 (0.45)		
Social cognition training				
present	15	0.52 (0.43)	0.46	0.64
absent	33	0.58 (0.42)		
Social skills education				
present	11	0.44 (0.29)	1.46	0.16
absent	37	0.60 (0.44)		
Attribution retraining				
present	3	0.32 (0.22)		
absent	45	0.58 (0.42)		
Other CBT component				
present	10	0.85 (0.52)		
absent	38	0.49 (0.36)		

toward very short term interventions in this sample, ranging from 15 minutes to 120 hours, with an average of nine hours of treatment(sd= 17.75). More than 50% of the treatments lasted six hours or less. This is significant because ten, 50-minute therapy sessions can be considered brief psychotherapy (Weakland, Fisch, Watzlawick, & Bodin, 1974). Therefore, many of the subjects in this sample received less treatment than what is considered brief therapy, although many of these children displayed clinically significant problem behavior. Length of therapy was not significantly correlated with effect size, r(45)= -0.11, p= 0.48. Similarily, the number of treatment sessions over the course of treatment ranged from one to 120, with the mean number of treatment sessions of 11.74 (sd= 17.56). Fifty percent of the studies offered 8 or fewer treatment sessions, again reflecting a restricted range. The number of treatment sessions was not significantly correlated with the effect size of treatment, r(47)= 0.19, p= 0.20. The possible interation between severity of a subjects symptom and the length of treatment relating to treatment effectiveness could not be explored further because of the restricted range of length of treatment.

The location of treatment was analyzed in relation to treatment effect. Most of the studies were conducted in schools (n= 30), two were conducted in a mental health clinic, twelve in an inpatient or residential setting, one at a camp, and three in more than one setting. Table 4 also reports the mean effect size based on treatment setting. The location of treatment was not significantly related to effect size, $F(4,45)= 1.68$, $p= 0.17$. Also, as presented in Table 4, there was no significant difference in treatment effectiveness for group therapies, individual therapies, or therapies which offered both treatment modalities, $F(2,47)= 0.62$, $p= 0.54$, one-tailed; nor was there a reliable difference when only individual and group treatments were compared, $T(43)= 0.80$, $p= 0.43$. The type of therapy offered was analyzed in relation to effect size. Based on a literature review of the CBT research with children, eight common components of treatment were identified: task-oriented problem-solving, social problem-solving, self-instruction training, role-play, concrete rewards either before or after success on a designated task, social cognition training, social skills training, and attribution retraining. Other treatment components were recorded as well, but were not seen frequently enough in

this sample to be included in the analysis.

Each treatment component was analyzed separately in relation to effect size by comparing the mean effect size for studies containing or not containing each component. As seen in Table 5, none of the treatment components on an individual basis were significantly related to treatment effectiveness. Is it possible that various combinations of treatment components were related to greater treatment effectiveness? It was impossible to examine the relationship between all combinations of components and effect size because there appeared to be no consistent combination of components. The treatment components were found in 40 unique combinations across the 63 comparisons in this meta-analysis and are presented in Table 6. For example, 25 comparisons used task oriented problem-solving, however this component was used in combination with other components in 17 different ways, as indicated in the bottom half of Table 6. These data suggest that CBT has not been consistently implemented in the literature.

Outcome Measure Characteristics

A separate meta-analysis was conducted using each outcome measure from the primary research studies, excluding measures from the study that produced an

Table 6

Combinations of Treatment Components

Components*									Frequency
1	2	3	4	5	6	7	8	9	
							1	0	1
						1	0	0	4
			1	0	0	0	1	0	4
			1	0	1	0	0	0	1
			1	0	1	0	1	0	2
			1	0	1	0	1	1	1
			1	0	1	1	0	0	1
		1	0	0	0	0	0	0	5
		1	0	0	0	0	0	1	2
		1	0	0	1	0	0	0	1
		1	0	0	1	1	0	0	2
		1	0	1	0	0	0	0	2
		1	1	0	0	0	0	0	3
		1	1	0	0	0	1	0	1
		1	1	0	0	1	0	0	1
		1	1	0	1	1	0	0	1
		1	1	1	0	0	0	0	1
		1	1	1	1	0	1	0	1
	1	0	0	0	0	0	0	0	1
	1	0	0	0	1	0	0	0	1
	1	0	1	0	1	0	1	0	1
	1	1	0	0	0	1	0	0	2
	1	1	1	0	0	0	0	0	1
1	0	0	0	0	0	0	0	0	2
1	0	0	0	0	0	1	0	0	1
1	0	0	1	0	0	0	0	0	1
1	0	1	0	0	0	0	0	0	2
1	0	1	0	1	1	0	0	0	3
1	1	0	0	0	0	0	0	0	1
1	1	0	0	1	0	0	0	0	1
1	1	0	1	0	0	0	0	0	1
1	1	0	1	0	1	0	1	0	1
1	1	0	1	1	0	0	0	0	1
1	1	0	1	1	1	1	0	0	1
1	1	1	0	0	0	0	0	0	2

Component coding:
1= task-oriented problem-solving
2= social problem-solving
3= self-instructions
4= role-playing
5= rewards
6= social cognition training
7= other CBT element
8= social skills training
9= attribution retraining

(continued)

Table 6 (continued)

Components*									Frequency
1	2	3	4	5	6	7	8	9	
1	1	1	0	0	0	0	1	0	1
1	1	1	0	0	0	1	0	0	2
1	1	1	0	1	0	0	0	0	1
1	1	1	0	1	1	0	0	1	1
1	1	1	1	1	1	0	0	0	3

*0= component not present in combination
1= component present in combination

outlier mean effect size. While 13 studies yielded more than one comparison, analyses were conducted based on measures from one comparison per study rather than from all comparisons, consistent with the other analyses. Two hundred and seventy-five separate outcome measures were coded based on 48 comparisons. The type and focus of the measure, the source of data, and the statistical method used to convert study statistics into an effect size were then examined in relation to effect size.

The statistical method used to estimate an effect size based on information in the study was significantly related to the resultant effect size, $F(7,274) = 8.24$, $p = 0.001$. Table 7 displays the average effect size related to method of effect size computation. Post hoc analysis with the Student-Newman-Keuls procedure at the 0.05 level revealed that the mean effect size based on methods 10 and 12 were significantly smaller than the mean effect size based on methods 1, 2, 3, 5, and 8.

Method 10 was an estimate of the effect size as zero when the measure was cited by the primary author but no posttest information was provided regarding the measure. This method of estimating an effect size was

Table 7

Mean Effect Size based on Method of Effect Size Calculation

Method of ES calculation	n	Mean ES (sd)
1. means, sd	180	0.52 (0.71)
2. ANOVA	10	0.87 (0.71)
3. t-test	7	0.96 (0.41)
4. raw data	0	---- ----
5. F test	11	0.91 (0.29)
6. ANCOVA	0	---- ----
7. Chi Squared	1	1.39
8. estimate from p	13	0.94 (0.35)
9. correlations	0	---- ----
10. estimated as 0	47	0.00 (0.00)
12. mixed methods	6	-0.003 (0.36)

expected to yield an effect size which was significantly smaller than the effect size estimates from other methods. Method 12 was used when a study reported using a measure but only reported the results of a few of the scales which compose the measure. In such instances the scale data provided were transformed into effect sizes, the effect sizes for the scales not reported were estimated as zero, and the all the effect sizes for the scales were averaged to yield one effect size estimate for the measure. Method 12 effect size estimates were nearly zero, therefore it is not suprising that they were also significantly smaller than effect size estimates from other methods.

Because of the significant impact on effect size of methods 10 and 12 for effect size computation on the resultant effect size calculations, analyses referred to as secondary analyses are reported in which measures using methods 10 and 12 in effect size estimation are removed. The purpose of the secondary analyses was to more carefully assess the impact that these two methods had on the outcome measure variables. (Analyses which include all measures used in the 48 comparisons will be referred to as the primary analysis). The secondary analysis to examine the relationship between type of

method used to estimate effect size and the resultant effect size revealed a marginally significant trend, $F(5,141) = 2.03$, $p = 0.08$.

Type of outcome measure was analyzed in relation to effect size. Outcome measures were classified as behavioral observation; sociometric or other type of peer rating; expert rating (e.g., therapist rating); standardized psychological measure (e.g., Kovac's (1981) Children's Depression Inventory) or a standardized behavioral checklist (e.g., Miller's (1972) School Behavior Checklist); an unstandardized psychological instrument or experimenter constructed instrument; an achievement test (e.g., Jastak, Bijou, and Jastak's (1980) Wide Range Achievement Test) or an intellectual measure (e.g., Weschler Intelligence Scale for Children-Revised, 1971); a cognitive measure (e.g., Kagan's (1966) Matching Familiar Figures Test) or other cognitive performance measure; or an objective performance measure (e.g., school grades, attendence record). The mean effect size based on the type of outcome measure is summarized in Table 8.

The type of measure used to collect data was significantly related to the effect size, $F(7,370) = 2.32$, $p = 0.02$. Post hoc analysis with

Table 8

Mean Effect Size for Type and Source of Outcome Measure

	n	Effect size (sd) Primary Analysis	n	Effect size (sd) Secondary Analysis
I. Type of Measure				
behavioral observation	59	0.49 (0.84)	45	0.63 (0.92)
peer rating/ sociometric	19	0.29 (0.32)	16	0.34 (0.32)
expert rating	11	1.12 (0.79)	11	1.12 (0.79)
normed psychological measure/checklist	36	0.41 (0.51)	28	0.55 (0.48)
unnormed psychological measure	70	0.44 (0.62)	53	0.58 (0.65)
achievement test/ intellectual test	11	0.63 (0.43)	10	0.69 (0.40)
cognitive/ performance measure	65	0.49 (0.64)	56	0.57 (0.65)
objective performance measure	4	0.37 (0.79)	3	0.49 (0.92)

(continued)

Table 8 (continued)

	n	Effect size (sd) Primary Analysis	n	Effect size (sd) Secondary Analysis
II. Source of Outcome Measure Information				
independent observer	50	0.51 (0.63)	37	0.68 (0.65)
parents	7	0.32 (0.33)	6	0.38 (0.33)
therapist	9	0.75 (0.62)	8	0.84 (0.99)
teacher/school	43	0.40 (0.62)	32	0.56 (0.63)
peer	19	0.29 (0.32)	16	0.34 (0.32)
self report	43	0.49 (0.67)	32	0.66 (0.71)
subject performance	97	0.56 (0.72)	85	0.64 (0.74)
other expert	7	-0.02 (0.42)	6	-0.03 (0.46)

Student-Newman-Keuls procedure indicated that the mean effect size of expert ratings was significantly larger than the mean effect size obtained from behavioral observations, peer ratings and sociometrics, standardized psychological measures and behavioral checklists, experimenter-constructed instruments, and cognitive or performance measures. Secondary analysis of the type of outcome measure and effect size revealed no significant effect, $F(7,141)= 1.50$, $p= 0.17$.

The source providing the outcome data was significantly related to effect size, $F(7,304)= 2.34$, $p= 0.02$. Table 8 summarizes the various sources of outcome measure information and the associated effect sizes. Post hoc analysis with the Student-Newman-Keuls procedure indicated the mean effect size from measures which obtained data from a therapist was significantly greater than the mean effect size from measures which obtained data from other expert judges. (Other expert sources of data included day and night workers and a head nurse in a residential setting, recreational staff, child care workers, camp counselors, and cottage leaders.) Secondary analysis indicated that the source of data had no significant influence on the effect size, $F(6,141)= 1.72$, $p= 0.12$.

Two general questions were asked of each measure: "Does this instrument measure change in cognitions or cognitive styles?"; and, "Does this instrument measure a change in overt behavioral adjustment?" Table 9 presents the mean effect size for measures based on answers to these questions. The mean effect size associated with measures of cognitive change was not significantly different from the mean effect size of measures which did not measure cognitive change in the primary analysis, $t(273) = -0.33$, $p = 0.74$, two-tailed, and in the secondary analysis, $t(211.81^*) = -0.08$, $p = 0.94$, two-tailed. Similarly, the effect size associated with measures of behavioral change was not significantly different from the effect size of measures which did not measure behavioral change in the primary analysis, $t(174.66^*) = 0.60$, $p = 0.55$, two-tailed, or in the secondary analysis, $t(129.69^*) = 0.20$, $p = 0.84$, two-tailed.

An assumption of cognitive-behavioral therapy is that changes in cognitive styles or cognitions translate into behavioral change (Kendall & Hollon, 1979; Meador & Ollendick, 1984; Pearl, 1985). Therefore, one would expect the mean effect size from measures of cognitive change and the mean effect size from measures of

Table 9

Mean Effect Size for Measures of Cognitive and Behavioral Change

	Primary Analysis		Secondary Analysis	
	Frequency	mean ES (sd)	Frequency	mean ES (sd)
I. Does this instrument measure change in cognitions or cognitive styles?				
Yes	137	0.50 (0.61)	112	0.60 (0.62)
No	138	0.47 (0.71)	110	0.59 (0.72)
II. Does this instrument measure a change in overt behavioral adjustment?				
Yes	100	0.45 (0.74)	78	0.58 (0.79)
No	175	0.50 (0.61)	144	0.60 (0.62)
III. Mean effect size for cognitive and behavioral change measures in studies which contained at least one measure of cognitive and one measure of behavioral change.				
a. Measures of cognitive change	29	0.59 (0.61)	22	0.67 (0.64)
b. Measures of behavioral change	29	0.33 (0.48)	22	0.49 (0.54)

behavioral change to be correlated in the 29 studies which contained at least one measure of cognitive change and at least one measure of behavioral change. However, the data did not support this hypothesis, $\underline{r}(29)= -0.20$, $\underline{p}= 0.30$. Secondary analysis also found no significant correlation between the mean effect size of measures of cognitive change and the mean effect size of measures of behavioral change within a study, $\underline{r}(22)= -0.28$, $\underline{p}= 0.21$. The mean effect size obtained in these latter analyses are presented in Table 9.

In summary, estimates of effect size as zero and estimates which are reduced to near zero because of missing data were significantly smaller than other estimates of effect size. As expected, no other method of effect size calculation was significantly related to the obtained effect size. Type of outcome measure and source of data were related to effect size, with expert ratings from therapists producing the largest estimates of effect size. The mean effect size from therapists was significantly greater the mean effect size from other types of experts such as nurses, hospital staff, and camp counselors. The mean effect size from expert ratings was significantly greater than the mean effect size from behavioral observations, peer ratings and

sociometrics, standardized psychological measures and behavior checklists, experimenter-constructed instruments, and cognitive or performance measures. The measures of cognitive change and behavioral change were not significantly correlated within a study, contrary to what was anticipated.

DISCUSSION

Cognitive behavioral therapy with children appears effective. The subjects exhibited a range of behavioral and cognitive problems and therapies varied in treatment focus, but the subjects in the treatment groups consistently showed greater improvement than the subjects in the control groups. Across 49 studies the mean treatment effect was 0.69 (sd= 0.94) and after removing one outlyer effect size, the mean treatment effect was 0.56 (sd= 0.42), indicating that subjects who received treatment advanced to the 75th and 71st percentile, respectively.

The direction and magnitude of the treatment effects in this study is consistant with other meta-analytic treatment reviews. For example, Casey and Berman (1985) reviewed 75 psychotherapy studies with children and found a treatment effect of 0.71 (sd= 0.73). The subset of 14 studies which utilized CBT yielded an effect size of 0.81 (sd=0.84). Duzinski (1987) found that children taught cognitive-behavioral strategies for educational purposes showed a treatment effect of 0.75. Smith and Glass (1977) reviewed 375 studies of psychotherapy with adults and report

treatment effectiveness of 0.68 (sd=0.67). Within Smith and Glass's (1977) sample of psychotherapy studies, the subset of 170 nonbehavioral studies which included cognitive therapies had a treatment effect of 0.60. Dush, Hirt, and Schroeder, (1983) conducted a meta-analysis of self-statement modification (SSM) with adults and found an effect size of 0.74 across all comparisons and 0.53 across comparisons which utilized a placebo control group rather than a no treatment control group. SSM is a principle technique of many cognitive-behavioral therapies and was defined as therapy which directly sought to modify covert self-statements (Dush et al., 1983).

There are several characteristics of this meta-analysis which strengthen confidence in the finding that CBT with children is effective. The literature search procedure was thorough, no evidence of bias was found between the published studies and unpublished dissertations, the design quality in the CBT studies was consistently high, the overall estimate of effect size was conservative, and the two meta-analyses provided valuable information both on design, treatment and subject characteristics, and on outcome measurement issues.

No evidence of publication bias was found in this sample of studies as the mean effect size of published studies and of unpublished dissertations were not significantly different. Ideally, the potential for publication bias would have been explored more thoroughly by collecting unpublished studies from conferences, writing to authors directly to request unpublished studies, and searching related fields for applicable unpublished literature. Glass and colleagues (1981) discuss instances in which publication bias was found in psychotherapy research, yet Dush and colleagues (1983) did not find evidence of publication bias in their meta-analysis of self-statement modification with adults. The conclusion that publication bias is not evident in this literature is bolstered by the knowledge that it would take an additional 86 studies with nonsignificant findings (e.g., effect size of 0.20) to reduce the effect size of 0.58 obtained from the 48 studies reviewed to nonsignificance.

The index of design quality provided evidence that the quality of the studies included in the meta-analysis was consistently high. On an index of zero to six, over half of the studies received a rating of five or six indicating they met all or all but one of the criteria

for good study design. None of the study design variables explored significantly influenced effect size. Generally, the studies were well designed; 40 of the 48 studies used randomized group assignment procedures or confirmed pretreatment equivalence. Further, 34 of the 48 comparisons utilized placebo-control groups. The placebo-control comparisons were expected to yield a more conservative estimate of treatment effectiveness by controlling for nonspecific and placebo related treatment gains. However, no difference was found between the effect size estimates of placebo-control group and no-treatment control group comparisons.

The estimate of ES in the present study is conservative because 53 of the 275 effect size calculations for outcome measures were based on an estimate of effect size as zero or near zero. These effect size estimates were used when data were either not reported or were incompletely reported for an outcome measure. Of the data reported (n=222), 17% produced an effect size of zero (n=3) or a negative effect size (n=35). It is unclear why data from 19% of the measures were not reported in the primary research. One suspects that a large proportion of the unreported data is statisitcally nonsignificant, however, based on

the percent of negative findings in the reported data, it is likely that only a relatively small percent of the unreported data is actually negative or equal to zero. If all the data had been available in the primary studies so that these zero estimates of effect size were unnecessary, it is likely that the additional nonsignificant positive findings would increase the average effect size. Another positive benefit to having data from all outcome measures is a more normal distribution of effect size because zero would not be overrepresented. Because meta-analytic technique relies on the availability of quantitative data in primary research, a straightforward remedy to this problem of zero and near zero estimates of effect size is more thorough reporting of primary research data, especially when the results are nonsignificant.

The type of outcome measure and the source of data had a significant effect on treatment effectiveness. The mean effect size of expert ratings was significantly greater than the mean effect size from behavioral observations, peer ratings and sociometrics, standardized pyschological measures and behavioral checklists, experimenter constructed instruments, and cognitive and performance measures. Specifically, the

role of the person considered the expert had a significant impact on the resultant effect size. When the source of an expert rating was a therapist, the mean effect size was 0.75 (sd= 0.97), while the expert ratings from a different source such as camp counselor, hospital staff, or child care worker yielded a mean effect size of -0.02 (sd= 0.42). In other words, the therapist's perceptions of subject improvement were significantly more positive than the perception of other people working with the children. Such apparent differences in perceptions of treatment change may be due to varying expectations for change, generalizability of change, duration of treatment effects, or quite simply self-serving therapist bias.

This meta-analysis took advantage of a range of information by analyzing the data in two ways: using one effect size per study, and using one effect size per outcome measure. This provided the opportunity to explore both study variables, such as design quality and treatment components, and outcome measure characteristics. Given the strong treatment effect size, lack of evidence for publication bias, and the high quality of studies, the next logical step was to explore variables which were expected to moderate

treatment effectiveness. Few significant relationships were found between design features and treatment effectiveness. The type of study design, group assignment procedure, pretreatment equivalence, and the number of days between the end of treatment and the collection of post-test data were not significantly related to treatment effectiveness. These nonsignificant findings seem to be due to the consistently high quality of research found in this sample of studies. The researchers are to be commended for this.

The length of treatment and the number of treatment sessions were two design features which were expected to impact upon treatment effectiveness. While conclusions from this review should be guarded because of a skew toward very short-term or brief treatments, neither length of treatment nor number of treatment sessions was significantly related to treatment effectiveness. These finding are counterintuitive, yet not novel (Casey & Berman, 1985; Dush et al., 1983; Duzinski, 1987). For example, Duzinski (1987) and Dush et al. (1983) found the total amount of time spent in therapy did not increase the size of the effects of treatment. Further, Duzinski (1987) found that treatment effects actually

declined with more treatment using problem-solving and modeling therapies. Casey and Berman (1985) found length of therapy negatively related to effect size but attributed this to the specificity of outcome measures for short- versus long-term therapies.

While length of therapy and the number of therapy sessions has long been an important clinical issue, mounting evidence casts doubt on the impact of treatment "dosage" on treatment effectiveness. However, it is important to consider the fact that while more than half the studies in this sample administered nine hours or less hours of treatment, ten, 50-minute sessions can be considered brief therapy (Weakland et al; 1974). Many of the subjects displayed clinically significant problems, but it was unclear if the dosage of therapy evident in research accurately reflects the dosage of CBT in clinics, mental health agencies, and private practitioner's offices. Therefore, the impact of length of treatment on treatment effectiveness may or may not generalize beyond the confines of controlled, short-term research interventions. Closer examination is warranted as a different, yet undisclosed variable may also explain the lack of impact that length of treatment seems to have on treatment outcome, and the findings

may or may not generalize to general clinical practice.

Few subject characteristics were significantly related to treatment effectiveness. Across all subjects, treatment effectiveness was not significantly related to the severity of the subject's problem. The specific type of target problem was not significantly related to treatment effect, nor was a broadband categorization of problems as either overcontrolled or undercontrolled (Achenbach, 1978). In other words, treatment was not more or less effective based on the targeted subject problem nor on the severity of the subject's problem. Consistent with Casey and Berman's (1985) findings, gender and subject age were not related to treatment effectiveness.

Also consistent with past findings, a nonlinear relationship between age and treatment effectivenesss was found (Duzinski, 1987). The studies were categorized according to subject age: 5 through 8 years, 9 to 11 years, and 11 years and older. The studies with children 9-11 years old (n=35) had the smallest effect sizes, and was significantly smaller than the effect size from the older children. It was expected that older children with more cognitive and verbal skills would benefit most from treatment, but it

was not expected that the 9-11 year old children would show the smallest treatment improvements. Upon examination, the 28 studies which had subjects in the middle age range did not differ from the overall sample of studies in design features, subject or treatment variables. But these studies accounted for 37 of the 53 outcome measure estimates of zero and near zero (effect size calculation methods 10 and 12). Therefore it is not clear if the lowered treatment effect for children ages 9 to 11 is actual or is an artifact created by lack of accurate reporting.

To examine the theoretical assumption that changes in cognitions or cognitive styles would translate into behavioral change (Kendall & Hollon, 1979), a subsample of studies which contained both at least one outcome measure of cogntive change and at least one outcome measure of behavioral change were examined (n=29). The mean effect sizes of cognitive change and of behavioral change were not significantly correlated, suggesting either the difficulty of defining and measuring cognitive change, or lack of support for a major premise of CBT.

Many reviewers have noted the importance of distinguishing between different types of CBT and the

critical elements which make CBT effective (Kendall, 1981; Lahey & Strauss, 1982; Urbaine & Kendall, 1980; Whalen, Henker, & Hinshaw, 1985; Wilson, 1984). A previous meta-analysis by Dush et al. (1983) examined the effects of treatment components in a meta-analysis of self-statement modification (SSM) therapy with adults. Dush and colleagues (1983) found the amount of variance accounted for by different CBT components varied based on the type of control group used in the primary research. Based on a multiple regression analysis, 10% of variability was accounted for by five treatment components in the no treatment control group comparisons. In the placebo control comparisons, 27% of the variability was accounted for by seven components. Conclusions from their findings must be guarded because of the small numbers of some treatment components and the discrepancies between components accounting for variance with no treatment control group comparisons and placebo control group comparisons. Dush et al's (1983) research suggested that a moderate amount of variance was explained by treatment components, but it did not reveal which components were essential to treatment or which combinations of treatment components were most effective.

The present review attempted to quantitatively examine different CBT techniques and combinations of components as they relate to treatment effectiveness. Distinctions were explored between different types of CBT by coding nine treatment components as present or absent in each treatment: task-oriented problem-solving, social problem-solving, self-instruction training, role-playing, concrete rewards, social cognition training, social skills training, attribution retraining, or "other" CBT components. The presence or absence of each individual component was analyzed in relation to treatment effectiveness, then the impact of various combinations of treatment components were examined. While CBT was clearly effective, the singular presence of none of the treatment components was significantly related to treatment effectiveness, indicating that none of the treatment components had a clear treatment advantage over other treatment components. Further, 40 different combinations of the nine therapy components were present in the 63 treatments.

The finding that the presence of no single component was significantly related to treatment effectiveness, as well as the plethora of combinations

of treatment components present in these studies, is perhaps the most important finding of this meta-analysis. The construct validity of the present use of the umbrella term "CBT" is brought into question by the diversity of treatment variations in this review. When researchers refer to CBT, they assume that others share a similar understanding of what CBT is, yet here the term CBT seems to defy definition. The sheer number of different combinations of treatment components in this study helps shed light on the extent to which cognitive behavioral therapies which are labeled the same may actually differ.

The question of construct validity regarding the term CBT is supported by the test of homogeneity of variance (Rosenthal & Rubin, 1982). While there is still debate concerning how to handle and interpret results when the independent studies included in the meta-analysis are heterogeneous (Wolf, 1986), substantial heterogeneity of variance was found in this sample of studies. When results are heterogeneous, the question arises as to whether or not the studies included in the meta-analysis have tested the same hypothesis. According to Wolf (1986), "Heterogeneity provides a warning that it may not be appropriate to

combine and synthesize all the study results in one meta-analysis" (p. 42). Finding substantial heterogeneity in the meta-analysis which was difficult to explain by subject, design, and treatment variables, and finding numerous combinations of treatment components, supports the arguments of those critical of applying meta-analytic technique to heterogeneous psychotherapy treatments. Again the "apples and oranges" issue is raised for CBT research with children. It seems that in this meta-analysis there may be 40 different apples and oranges.

The treatments reviewed in this meta-analysis were clearly effective in improving children's cognitive and behavioral functioning. This finding is durable, as it would take an additional 86 studies with nonsignificant findings (ES= .20) to reduce the effect size of 0.58 obtained from these 48 studies to nonsignificance. Further, it seems that researchers have may have successfully modified theoretical concepts to fit specific problems, populations, and developmental levels of children. But the lack of significant differences in treatment effectiveness based on the presence or absense of each treatment component, the large number of different combinations of treatment components, the

heterogeneity of variance in effect size, and the lack of correlation within studies between outcome measures of cognitive change and behavioral change, lead to the conclusion that while CBT is clearly an effective form of treatment, we are no closer to knowing the answer to two basic questions: (1) What exactly is included in the rubric of CBT? and (2) What critical elements of treatment make CBT effective?

REFERENCES

Abikoff, H. (1978). Cognitive training interventions in children: Review of a new approach. Journal of Learning Disabilities, 12, 65-77.

Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. Psychological Bulletin, 85, 1275-1301.

Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method. Psychological Bulletin, 99, 388-399.

Brown, R. T., Wynne, M. E., & Medenis, R. (1985). Methylphenidate and cognitive therapy: A comparison of treatment approaches with hyperactive boys. Journal of Abnormal Child Psychology, 13, 69-87.

Bryant, F. B. & Wortman, P. M. (1984). Methodological issues in the meta-analysis of quasi-experiments. In W. H. Yeaton, & P. M. Wortman (Eds.). Issues in data synthesis. New Directions for Program Evaluation, no. 24. San Fransisco: Jossey-Bass, December 1984.

Bullock, R. J. & Svyantek, D. J. (1985). Analyzing meta-analysis: Potential problems, and unsuccessful replication, and evaluation criteria. Journal of Applied Psychology, 70, 108-115.

Camp, B. W., Blom, G. E., Herbert, F., & van Doorninck, W. J. (1977). "Think aloud": A program for developing self-control in young aggressive boys. Journal of Abnormal Child Psychology, 5, 157-169.

Cannavo-Antognini, F. (1979). Cognitive modeling of aggressive behavior in children: A study of setting generality of impulsive control. Dissertation Abstracts International, 40, 3384B.

Casey, R. J., & Berman, J. S. (1985). The outcome of psychotherapy with children. Psychological Bulletin, 98, 388-400.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. Journal of Abnormal and Social Psychology, 65, 145-153.

Cohen, J. (1977). Statistical Power Analysis for the Behavioral Sciences (Rev. ed.) New York: Academic Press.

Cole, P. M. & Kazdin, A. E. (1980). Critical issues in self-instruction training with children. Child Behavior Therapy, 2, 1-21.

Cook,T. D. & Leviton, L. C. (1980). Reviewing the literature: A comparison of traditional methods with meta-analysis. Journal of Personality, 48, 449-472.

Dush, D. M., Hirt, M. L., & Schroeder, H. (1983). Self-statement modification with adults: A meta-analysis. Psychological Bulletin,94, 408-422.

Duzinski, G. A. (1987). The educational utility of cognitive behavior modification strategies with children: A quantitative synthesis. Dissertation Abstracts International, 48, 339A.

Dweck, C.S. (1975). The role of expectations and attributions in the alleviation of learned helplessness. Journal of Personality and Social Psychology, 31, 674-685.

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

Ellis, A. (1963). Reason and Emotion in Psychotherapy. NY: Stewart.

Eyseneck, H. J. (1978). An exercise in mega-silliness. American Psychologist, 33, 517.

Fiske, D. W. (1983). The meta-analytic revolution in outcome research. Journal of Consulting and Clinical Psychology,51, 65-70.

Glass, G. V. (1978). Integrating findings: The meta-analysis of research. In L. S. Shulman (Ed.), Review of Research in Education. vol. 5. Itasca, IL: Peacock Publishers. (pp. 351-379).

Glass, G. V., McGaw, B. & Smith, M. L. (1981). Meta-analysis in Social Research. Beverly Hills, CA: Sage.

Gresham, F. M. (1985). Utility of cognitive-behavioral procedures for social skills training with children: A critical review. Journal of Abnormal Child Psychology, 13, 411-423.

Gresham, F. M. (1981). Social skills training with handicapped children: A review. Review of Educational Research, 51, 139-176.

Gresham, F. M. & Lemanek, K. L. (1983). Social skills: A review of cognitive-behavioral training procedures with children. Journal of Applied Developmental Psychology, 4, 239-261.

Gresham, F. M., & Nagle, R. J. (1980). Social skills training with children: Responsiveness to modeling and coaching as a function of peer orientation. Journal of Consulting and Clinical Psychology, 48, 718-729.

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. Psychological Bulletin,92, 490-499.

Holmes, C. T. (1984). Effect size estimation in meta-analysis. Journal of Experimental Education, 52, 106-109.

Hobbs, S.S., Moguin, L.E., Tyroler, M., & Lahey, B.B. (1980). Cognitive-behavioral therapy with children: has clinical utility been demonstrated? Psychological Bulletin, 87, 147-165.

Hunter, P. A. (1985). Evaluation of a cognitive behavior therapy program with aggressive male children in residential treatment. Dissertation Abstracts International, 46, 2811B.

Kagan, J. (1966). Reflection-impulsivity: The geneerality and dynamics of conceptual tempo. Journal of Abnormal Psychology, 71, 17-24.

Kahl, B. (1985). A school-based cognitive-behavior modification program for hyperaggressive children. Dissertation Abstracts International, 45, 126A.

Jastak, J. F., Bijou, S. W., & Jastak, S. R. (1965). Wide Range Achievement Test. Wilmington, DE: Guidance Associates.

Kendall, P.C. (1981). Cognitive-behavioral interventions with children. In B.B. Lahen and A. E. Kazdin (Eds.), Advances in Clinical Psychology (vol. 4). NY: Plenum Press. (pp. 53-90).

Kendall, P. C., & Braswell, L. (1982). Cognitive-behavioral self-control therapy for children: A components analysis. Journal of Consulting and Clinical Psychology, s50, 672-689.

Kendall, P. C. & Hollon, S. D., (Eds.). (1979). Cognitive-behavioral Interventions: Theory, Research, and Procedures. NY: Academic Press.

Kovacs, M. (1978). Children's Depression Inventory. Pittsburg, PA: University of Pittsburgh Press.

La Greca, A. M., & Santogross, D. A. (1980). Social skills training with elementary school students: A behavioral group approach. Journal of Consulting and Clinical Psychology, 48, 220-227.

Landman, J. T., & Dawes, R. M. (1982). Psychotherapy outcome: Smith and Glass' conclusions stand up under scrutiny. American Psychologist, 37, 504-516.

Lahey, B.B. & Strauss, C.C. (1982). Some considerations in evaluating the clinical utility

of cognitive-behavioral therapy with children. School Psychology Review,11, 67-74.

Light, R. J. & Pillemer, D. B. (1984). Summing up: The Science of Reviewing Research. Boston: Harvard Press.

Light, R. J. & Smith, P. V. (1971). Accumulating evidence: Procedure for resolving contradictions among different research studies. Harvard Educational Review, 41, 429-471.

Mansfield, R. S. & Busse, T. V. (1977). Meta-analysis of research: A rejoinder to Glass. Educational Researcher, 6, 3.

Meador, A. E., & Ollendick, T. H. (1984). Cognitive behavior therapy with children: An evaluation of its efficacy and clinical utility. Child and Family Behavior Therapy, 6, 25-44.

Miller, L. C. (1972). School behavior checklist: An inventory of deviant behavior for elementary school children. Journal of Consulting and Clinical Psychology, 38, 134-144.

Orwin, R. (1983). A fail-safe N for effect size. Journal of Educational Statistics, 8, 157-159.

Meichenbaum, D. H., & Goodman, J. (1971). Training impulsive children to talk to themselves: A means

of developing self-control. Journal of Abnormal Psychology, 77, 115-126.

Meichenbaum, D. H. (1977). Cognitive-behavior modification: An Integrative Approach. NY: Plenum Press.

Meador, A. E. & Ollendick, T. H. (1984). Cognitive-behavior therapy with children: An evaluation of its efficacy and clinical utility. Child and Family Behavior Therapy, 6, 25-44.

Pearl, R. (1985). Cognitive-behavioral interventions for increasing motivation. Journal of Abnormal Child Psychology, 13, 443-454.

Presby, S. (1978). Overly broas categories obscure differences in therapies. American Psychologist, 33, 514-515.

Presley, M. (1979). Increasing childrens self-control through cognitive interventions. Review of Education Research, 49, 319-370.

Rosenthal, R., and Rubin, D. (1982). Further meta-analytic procedures for assessing cognitive gendeer differences. Journal of Educational Psychology, 74, 708-712.

Smith, M. L. (1980). Publication bias and meta-analysis. Evaluation in Education, 4, 22-24.

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

Spivack, G. & Shure, M. B. (1974). Social Adjustment of Young Children, a Cognitive Approach to solving real-life problems. San Fransisco: Jossey-Bass.

Thomas, A., & Pashley, B. (1982). Effects of classroom training on learning disabled students' task persistence and attributions. Learning Disability Quarterly, 5, 133-144.

Urbain, E. S. & Kendall, P. C. (1980). Review of social-cognitive problem-solving interventions with children. Psychological Bulletin, 88, 109-143.

Weakland, J. H., Fisch, R., Watzlawick, P., & Bodin, A. M. (1974). Brief therapy: Focused problem resolution. Family Process, 2, 141-166.

Whalen, C. K., Henker, B., & Hinshaw (1985). Cognitive-behavioral therapies for hyperactive children: Premises, problems, and prospects. Journal of Abnormal Child Psychology, 13, 391-410.

Wilson, R. (1984). A review of self-control treatments for aggressive behavior. Behavioral Disorders, 9, 131-140.

Wolf, F. M. (1986). Meta-analysis. Quantitative Methods for Research Synthesis. Beverly Hills, CA: Sage.

Wortman, P. M. & Bryant, F. B. (1985). School segregation and black achievement. Sociological Methods and Research, 13, 289-324.

APPENDIX A

Studies Included in Meta-Analysis

Braswell, L. (1984). Cognitive-behavioral therapy with an inner city sample of non-self-controlled children. Dissertation Abstracts International, 45, 2679B.

Brown, R. T. (1980). Impulsivity and psychoeducational intervention in hyperactive children. Journal of Learning Disabilities, 13, 19-24.

Brown, R. T., Wynne, M. E., & Medenis, R. (1985). Methylphenidate and cognitive therapy: A comparison of treatment approaches with hyperactive boys. Journal of Abnormal Child Psychology, 13, 69-87.

Camp, B. W., Blom, G. E., Herbert, F., Van Doorninck, W. J. (1977). "Think aloud": A program for developing self-control in young aggressive boys. Journal of Abnormal Child Psychology, 5, 157-169.

Cannavo-Antognini, F. (1979). Cognitive modeling of aggressive behavior in children: A study in setting generality of impulsive control. Dissertation Abstracts International, 40, 3384B.

Cotugno, A. J. (1982). Cognitive control methods in the treatment of special education children. Psychology in the Schools, 19, 517-525.

Craig, J. L. (1984). The influence of relaxation and cognitive training on the behavior, academic performance, perseverance, and problem-solving of hyperactive children. Dissertation Abstracts International, 45, 1340A.

Finch, A. J., Wilkinson, M. D., Nelson, W. M., Montgomery, L. E. (1975). Modification of impulsive cognitive tempo in emotionally disturbed boys. Journal of Abnormal Child Psychology, 13, 49-52.

Fleming, C. C. (1982). Evaluation of an anger management program with aggressive children in esidential treatment. Dissertation Abstracts International,43, 4143B.

Fowler, J. W., & Peterson, P. L. (1981). Increasing reading persistence and altering attributional stype of learned helpless children. Journal of Educational Psychology, 73, 251-260.

Gorton, M. V. (1985). The efficacy of teaching soical skills to socially isolated preadolescents. Dissertation Abstracts International, 47, 4328B. (University Microfilms No. ADG86-22832)

Gresham, F. M., & Nagle, R. J. (1980). Social skills training with children: Responsiveness to modeling

and coaching as a function of peer orientation. Journal of Consulting and Clinical Psychology, 48, 718-729.

Guralnick. M. (1976). Solving complex discrimination problems: Techniques for the development of problem-solving strategies. American Journal of Mental Deficiency, 81, 18-25.

Harris, K. R., & Brown, R. D. (1982). Cognitive-behavior modification and informed teacher treatments for shy children. Journal of Experimental Education, 50, 137-143.

Hunter, P. A. (1985). Evaluation of a cognitive behavior therapy program with aggressive male children in residential treatment. Dissertation Abstracts International, 46, 2811B.

Jasnow, M. (1982). Effects of relaxation training and rational emotive therapy on anxiety reduction in sixth grade children. Dissertation Abstracts International, 43, 4149B.

Kahl, B. (1985). A school-based cognitive-behavior modification program for hyperaggressive children. Dissertation Abstracts International, 45, 126A. (University Microfilms No. ADG84-09259)

Kanfer, F. H., Karoly, P., & Newman, A. (1975). Reduction of children's fear fo the dark by competencrelated and situational threat-related verbal cues. Journal of Consulting and Clinical Psychology, 43, 251-258.

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

Kendall, P. C., & Braswell, L. (1982). Cognitive-behavioral self-control therapy for children: A components analysis. Journal of Consulting and Clinical Psychology, 50, 672-689.

Kendall, P. C., & FInch, A. J. (1979). Analysis of changes in verbal behavior following a cognitive-behavioral treatment for impulsivity. Journal of Abnormal Child Psychology, 7, 455-463.

Kendall, P. C., & FInch, A. J. (1978). A cognitive-behavioral treatment for impulsivity: A group comparison study. Journal of Consulting and Clinical Psychology, 46, 110-118.

Kendall, P. C., & Wilcox, L. E. (1980). Cognitive-behavioral treatment for impulsivity:

Concrete versus conceptual training in non-self-controlled problem children. Journal of Consulting and Clinical Psychology, 48, 80-91.

Kendall, P. C., & Zupan, B. A. (1981). Individual versus group application of cognitive-behavioral self-control procedures with children. Behavior Therapy, 12, 344-359.

Kettlewell, P. W., & Kausch, D. F. (1983). The generalization of the effects of a cognitive-behavioral treatment program for aggressive children. Journal of Abnormal Child Psychology, 11, 101-114.

Kingsley, G. J. (1987). The efficacy of a self-instruction attribution retraining program in alleviating learned helplessness and depressive symptomotology in learning disabled children. Dissertation Abstracts International, 48, 879B.

Ladd, G. W. (1981). Effectiveness of a social learning method for enhancing children's social interaction and peer acceptance. Child Development, 52, 171-178.

La Greca, A. M., & Santogross, D. A. (1980). Social skills training with elementary school students: A behavioral group approach. Journal of Consulting and Clinical Psychology, 48, 220-227.

Mannarino, A. P., Christy, M., Durlak, J. A., Mognussen, M. G. (1982). Evaluation of social competence training in the schools. Journal of School Psychology, 20, 11-19.

McDonoush, T. J. (1979). A comparison of self-instruction and play group therapy in lessening impulsivity and subsequent hyperaggressivity in fourth, fifth, and sixth grade boys. Dissertation Abstracts International,40, 4493B.

Meichenbaum, D. H., & Goodman, J. (1971). Training impulsive children to talk to themselves: A means of developing self-control. Journal of Abnormal Psychology, 77, 115-126.

Moore, S. F., & COle, S. O. (1978) Cognitive self-mediation training with hyperkinetic children. Bulletin of the Psychonomic Society, 12, 18-20.

Oden, S., & Asher, S. R. (1977). Coaching children in social skills for friendship making. Child Development, 48, 495-506.

Olexa, D. F., & Forman, S. G. (1984). Effects of social problem-solving training on classroom behavior of urban disadvantaged students. Journal of School Psychology, 22, 165-175.

Parrish, J. M., & Erickson, M. T. (1981). A comparison of cognitive strategies in modifying the cognitive style of impulsive third-grade children. Cognitive Therapy and Research, 5, 71-84.

Reid, M. K., & Borkowski, J. G. (1987). Causal attributions of hyperactive children: Implications for teaching strategies and self-control. Journal of Educational Psychology, 79, 296-307.

Roseby, V., & Deutsch, R. (1985). Children of seperation and divorce: Effects of a social role-taking group intervention on fourth and fifth-graders. Journal of Child Clinical Psychology, 14, 55-60.

Rotheram, M. J. (1982). Social skills training with underachievers, disruptive, and exceptional children. Psychology in the Schools, 19, 532-539.

Sackles, J. A. (1980). An evaluation of three treatment programs for anger control in young adolescents. Dissertation Abstracts International, 42, 1189B.

Small, R. W., & Schnike, S. P. (1983). Teaching competence in residential group care: Cognitive problem-solving and interpersonal skills training with emotionally disturbed preadolescents. Journal of Social Service Research, 7, 1-16.

Smith, J. M. (1980). The effects of problem-solving groups on the cognitive style dimension of reflection-impulsivity in delinquent youths. Dissertation Abstracts International, 42, 149A.

Stark, K. D., Reynolds, W. M., & Kaslow, N. J. (1987). A comparison of the relative efficacy of self-control therapy and a behavioral problem-solving therapy for depression. Journal of Abnormal Child Psychology, 15, 91-113.

Stefanek, M. E. (1984). Self-statement utilization and social skills training with elementary school-aged children. Dissertation Abstracts International, 46, 663B.

Terwilliger, C. L. (1987). Teaching problem solving strategies to children with attention deficit disorder. Dissertation Abstracts International,48, 908A.

Thomas, A., & Pashley, B. (1982). Effects of classroom training on learning disabled students' task persistence and attributions. Learning Disabilities Quarterly, 5, 133-144.

Wen, S. (1974). Modifying children's impulsivity through redundancy and variability training. Psychology in the Schools, 11, 440-444.

Wolfson, J. (1982). Self-talk and exposure in the reduction of avoidance in aquaphobic children. Dissertation Abstracts International,43, 2721B.

Zelie, K., Stone, C. I., & Lehr, E. (1980). Cognitive-behavioral intervention in school discipline: A preliminary study. Personnel and Guidance Journal, 59, 80-83.

Zitomer, E. A. (1981). Decreasing impulsivity in grade school children: Self-instructional training and imagry. Dissertation Abstracts International,42, 2093B.

APPENDIX B

Studies Meeting Inclusionary Criteria but Excluded from Meta-Analysis

Arnold, S. C., & Forehand, R. (1978). A comparison of cognitive training and response cost procedures in modifying cognitive style of impulsive children. Cogntitive Therapy and Research, 2, 183-187.

Berler, E. S., Gross, A. M., Drabman, R. J. (1982). Social skills training with children: Proceed with caution. Journal of Applied Behavior Analysis, 15, 41-53.

Bierman, K. L., & Furman, W. (1984). The effects of social skills training and peer involvement on the social adjustment of preadolescents. Child Development, 55, 151-162.

Bornstein, M. R., Bellack, A. S., & Hersen, M. (1977). Social skills training for unassertive children: A multiple baseline analysis. Journal of Applied Behavior Analysis, 10, 183-195.

Cohen, N. J., Sullivan, J., Minde, K., Novak, C., & Helwig, C. (1981). Evaluation of the relative effectiveness of methylphenidate and cognitive behavior modification in the treatment of kindergarden-aged hyperactive children. Journal of Abnormal Child Psychology, 9, 43-54.

Combs, M. L., Lahey, B. B. (1981). A cognitive social skills training program, evaluation with young children. Behavior Modification, 5, 39-60.

Cooke, T. P., Appolloni, T. (9176). Developing positive social-emotional behaviors: A study of training and generalization effects. Journal of Applied Behavior Analysis, 9, 65-78.

Eastman, B. G., & Rasbury, W. C. (1981). Cognitive self-instruction for the control of impulsive classroom behavior: Ensuring the treatment package. Journal of Abnormal Child Psychology, 9, 381-387.

Foreman, S. G. (1980). A comparison of cognitive training and response cost procedures in modifying aggressive behavior of elementary school children. Behavior Therapy, 11, 594-600.

Holmes, C. S. (1981). Reflective training and causal attributions in impulsive mildly retarded children. Journal of Clinical Child Psychology, 10, 194-199.

Litrownik, A. J., Freitas, J. L., & Franzini, L. R. (1978). Self-regulation in mentally retarded children: Assessment and training of self-monitering skills. American Journal of Mental Deficiency, 82, 499-506.

Lochman, J. E., Burch, P. R., Curry, J. F., & Lampron, L. B. (1984). Treatment and generalization effects of cognitive-behavioral and goal-setting interventions with aggressive boys. Journal of Consulting and Clinical Psychology, 52, 915-916.

Nelson, W. J., Birkimer, J. C. (1978). Role of self-instruction and self-reinforcement in the modification of impulsivity. Journal of Consulting and Clinical Psychology, 46, 183.

Zahavi, S., & Asher, S. R. (1978). The effect of verbal instructions on preschool children's aggressive behavior. Journal of School Psychology, 16, 146-153.

APPENDIX C

Coding Sheet--CBT with Children

general code 99= unknown/unclear

I. GENERAL INFORMATION

1 _____ study ID number
author, year
title
source
2 _____ 1=published 2=dissertation
3 _____ year of publication (last 2 digits only)
5 _____ total number of comparisons
7 _____ total number of outcome measures at post-intervention

II. DESIGN FEATURES

23 _____ type of design
1= pretest-poesttest with nonequivalent control group (NECG)
2= posttest only with NECG
3= randomized true experiment
4= other (e.g., matching)
24 _____ group assignment procedure (01-05)
1= random
2= matching
3= available intact
4= voluntary self-selection
5=other
26 _____ number of days post-Tx measures made following Tx
27 _____ is the number of days following Tx that measures taken an estimate 0=no 1=yes
28 _____ follow-up information? 0=no 1=yes
29 _____ length FU in days
30 _____ same measures used at FU and post-Tx? 0=no 1=yes
31 _____ sample size LE 30 =0 sample size GT 30= 1
32 _____ random assignment or pretreatment equivalence 0=no 1=yes
33 _____ attrition less than 10% or equal 0=no 1=yes
34 _____ at least one normed measure or blinded behavioral outcome measure 0=no 1=yes
35 _____ type control group 0= no TX cntl 1= placebo cntl

36 _____ same instruments at pre- and post-test 0=no 1=yes
37 _____ sum of coding items 31 - 36

III. SUBJECT INFORMATION

38 _____ mean age
39 _____ is mean age an estimate? 0=no 1=yes
40 _____ number of subjects in study
41 _____ number of male subjects
42 _____ number of subjects in treatment group
43 _____ number of subjects in control/comparison group
44 _____ ethnic sample characteristics
1= majority or all white
2= majority or all minority
3= mixed
45 _____source of subjects
1=inpatients/residential
2=outpatients
3=volunteers for special project
4=chosen through problem-orientedobservation, measurement, or recommendation
46 _____ LD problem? 0= no 1= yes 3= present in some, not all Ss
46.1_____ Is it reasonable to consider the subject's problem a clinically significant problem? 0= no 1= certainly yes
2= uncertain or unclear (maybe, if...)
47 _____ target problem
01= social isolate
02= fear/phobias
03= anxiety
04= enuresis
05= somatic problems
06= depression
07= other internalizing symptomotology or mix of 1-6
08= impulsivity/hyperactivity
09= non-compliant/management problem/behavior problem
10= psychotic/autistic
11= other externalizing symptomotology or mix of 8 - 10
12= socialskills,undefined
13= mix of 1-12 further symptom description:

IV. TREATMENT CHARACTERISTICS

48 _____ treatment modality 1=individual 2=group 3=mixed
49 _____ length of treatment in hours
50 _____ length of treatment estimated? 0=no 1=yes
51 _____ number of treatment sessions
52 _____ number of sessions estimated? 0=no 1=yes
53 _____ treatment setting
1= school
2= home
3= mental health, CMHC, psyc clinic
4= general hospital or dental clinic
5= residential treatment (psychiatric or special school)
6= camp
7= combo of at least two of the above
8= other

components of treatment: 0=absent 1=present
57 _____ problem-solving, task oriented
57.1 _____ social problem-solving
58 _____ self instructions
60 _____ role-play
62 _____ concrete rewards (before or after)
65 _____ social cognition (affective ed, perspective taking)
67 _____ other:
67.1 _____ social skills training
67.2 _____ attribution retraining

V. COMPARISON/ EFFECT SIZE INFORMATION

68 _____ type of comparison 1= Tx vs placebo control 2= Tx vs no treatment control
69 _____ sample size in ES calculation
70 _____ mean ES at post-treatment
71 _____ w= 2N/ 8+d(squared)
72 _____ wd

VI. OUTCOME MEASURES

name of measure:

73 _____ type of measure:

1= behavioral observation
2= peer rating/sociometric
3= "expert" rating or behavioral checklist
4= normed psychological measure
5= nonnormative or experimentor constructed instrument
6= achievement test or intellectual measure
7= cognitive measure or other performance measure
8= objective performance measure (days in school, arrests)
9= other

74 _____ source of data:

1= independent observers
2= parents
3= therapist
4= teacher/school
5= peer
6= subject self report
7= subject performance measure (on achvt,IQ, or cog measure)
8= other(expert judges not 1-7)
9= mixed

75 _____ dimension of adjustment:

01= fear/anxiety
02= cognitive skills
03= global adjustment
04= social adjustment/social skills
05= achievement
06= personality
07= self-esteem
08= bed-wetting
09= mixed

76 _____ Does this instrument measure change in cognitions or thought process that are the focus of treatment? 0=no 1=yes

77 _____ Does this instrument measure a change in overt behavior that is the focus of treatment 0=no 1=yes

79 _____ method used in calculation of ES:
01= means, standard deviations
02= ANOVA (Holmes, 1986)
03= t test
04= raw data
05= F test
06= ANCOVA
07= Chi*2/nonparametric
08= estimate from p
09= correlations
10= effect size estimated as zero
12= mixed methods
80 _____ ES at post-treatment

APPROVAL SHEET

The thesis submitted by Teresa Suriano Fuhrman has been read and approved by the following committee:

Dr. Joseph A. Durlak, Director
Professor, Clinical Psychology, Loyola

Dr. Fred B. Bryant
Asssociate Professor, Social Psychology, Loyola

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorperated and that the thesis is now given final approval by the Committee with reference to content and form.

The thesis is therefore accepted in partial fulfillment of the requirements for the degree of Master of Arts.

Nov 2, 1988
Date

Joseph A Durlak
Director's Signature