A Meta-Analysis of the Treatment of Internalizing Disorders in Children and Adolescents

Kathryn Ann McGlinchey
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_theses

Part of the Psychology Commons

Recommended Citation
https://ecommons.luc.edu/luc_theses/4187

This Thesis is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Master's Theses by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 1996 Kathryn Ann McGlinchey
LOYOLA UNIVERSITY CHICAGO

A META-ANALYSIS OF THE TREATMENT OF INTERNALIZING DISORDERS IN CHILDREN AND ADOLESCENTS

A THESIS SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
MASTER OF ARTS
DEPARTMENT OF PSYCHOLOGY

BY
KATHRYN ANN MCCLINCHEY

CHICAGO, ILLINOIS
MAY, 1996
Copyright by Kathryn Ann McGlinchey, 1996
All rights reserved.
ACKNOWLEDGMENTS

I am very pleased to express my gratitude to those people who were involved either directly or indirectly in the preparation of my thesis. In particular, I would like to recognize my thesis committee for guiding this project through to its completion. A special thanks goes to the chair of my committee, Dr. Joseph Durlak. Over the past two years, he has been a continual source of support and encouragement to me. His knowledge and enthusiasm for the subject matter and for the techniques of meta-analysis has enabled me to not only complete this project but to also consider it a valuable and enjoyable learning experience. I would also like to thank Dr. Paul Jose for his helpful comments and his willingness to contribute to this project.

Additionally, I wish to recognize others who were instrumental in this process as well. My most sincere gratitude goes to my family and friends for the many ways in which they inspired and encouraged me. In particular, I want to thank my parents who have unconditionally supported all of my endeavors, both academic and personal. Finally, I am very appreciative and indebted to my friends who have offered me both humor and perspective throughout the duration of this project, making its completion possible.
# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** ......................................................... iii

**LIST OF TABLES** ............................................................ v

Chapter

I. **OVERVIEW** ................................................................. 1

II. **LITERATURE REVIEW** .................................................. 4

   - Introduction to Child Psychotherapy Research .................. 4
   - Rationale for Current Investigation ............................. 20
   - Hypotheses ............................................................... 22

III. **METHOD** ................................................................. 28

   - Literature Search .................................................... 28
   - Coding Procedures .................................................. 29
   - Analysis of Interventions ......................................... 30

IV. **RESULTS** ................................................................. 32

   - Sample Characteristics ............................................ 32
   - Treatment Characteristics ........................................ 35
   - Methodological Features .......................................... 38
   - Homogeneity of Effect Sizes ...................................... 42
   - Treatment Outcomes ............................................... 43
   - Multiple Regression Analyses .................................... 57

V. **DISCUSSION** ............................................................... 60

   - Efficacy of Treatments ............................................ 60
   - Discussion of Descriptive Findings ............................. 69
   - Limitations of the Current Review .............................. 74
   - Future Directions .................................................... 75

**APPENDIX**

A. **LIST OF JOURNALS USED IN MANUAL SEARCH PROCEDURES** ........... 79

B. **CODING SCHEMA FOR META-ANALYSIS OF INTERNALIZING DISORDERS** ... 81

**REFERENCES** ................................................................. 88

**VITA** ................................................................. 92
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sample Characteristics of Reviewed Studies</td>
<td>33</td>
</tr>
<tr>
<td>2. Treatment Characteristics of Reviewed Studies</td>
<td>36</td>
</tr>
<tr>
<td>3. Methodological Characteristics of Reviewed Studies</td>
<td>39</td>
</tr>
<tr>
<td>4. Findings from Homogeneity Analyses for Clinical Variables</td>
<td>45</td>
</tr>
<tr>
<td>5. Findings from Homogeneity Analyses for Interactions</td>
<td>54</td>
</tr>
<tr>
<td>6. Findings from Homogeneity Analyses for Methodological Variables</td>
<td>56</td>
</tr>
<tr>
<td>7. Results from Regression Analyses</td>
<td>59</td>
</tr>
</tbody>
</table>
CHAPTER I

OVERVIEW

Internalizing disorders are quite common among children and adolescents and may persist over time if left untreated (Ollendick & King, 1994). Despite their high prevalence rates, little is known about the treatment of internalizing problems since research on child treatment has tended to favor the study of externalizing problems. For example, many of the meta-analyses of child psychotherapy have included a preponderance of studies focusing on externalizing rather than internalizing symptomatology (Weisz, Weiss, Alicke, & Klotz, 1987; Weisz, Weiss, Han, Granger, & Morton, 1995).

Therefore, the findings from current evaluations of child psychotherapy may not generalize to internalizing disorders. As a result, there is a need to examine current treatments for children with internalizing disorders. More specifically, the intent of this proposed study is to assess the effectiveness of treatments for internalizing disorders and to identify factors that influence treatment outcome.

This review will begin by discussing the nature and symptomatology of various internalizing disorders. Following this section, attention will be devoted to summarizing the findings from the major meta-analytic
reviews of child and adolescent psychotherapy. Potential treatment outcome moderators which have been identified by these reviews will be discussed individually. Their possible influence on the treatment of internalizing disorders will be considered as well. Finally, the rationale for a meta-analytic review of internalizing disorders will be presented along with the hypotheses for this investigation.

**Nature of Internalizing Disorders**

Internalizing disorders include depression, social isolation and withdrawal, anxiety disorders, and psychosomatic disorders. As a class of disorders, internalizing problems are known for their inner-directed focus in which the primary symptoms are associated with over-controlled behaviors (Reynolds, 1992). The distinction between internalizing and externalizing disorders emerged largely from the empirical work conducted by Achenbach's research lab (Achenbach & Edelbrock, 1978; Achenbach & McConaughy, 1987). Although some discrepancies and overlap do exist, externalizing disorders are commonly grouped and referred to as behavior disorders while internalizing disorders are more often known and referred to as emotional disorders (Reynolds, 1992).

The most basic feature of internalizing disorders is the subjective feelings of distress experienced by the child or adolescent (Reynolds, 1992). While these disorders may
differ in the specific type of experienced internalized distress, they are similar in that the symptoms are not easily detected. Given the inner-experienced or subjective nature of the distress, identification of these symptoms can be difficult (Reynolds, 1992). In addition, the covert nature of internalizing symptomatology may present unique challenges for treatment. However, an understanding of the specific signs or features of internalizing disorders can facilitate their identification and eventual treatment.

This section will introduce the findings from prior meta-analyses that have attempted to evaluate the effectiveness of child psychotherapy. The treatment moderators believed to hold particular significance for the treatment of internalizing disorders will be presented first. These variables include comorbidity, gender and age issues, types of treatment, types of therapist, and outcome measures. This will be followed by a discussion of additional variables which have emerged as possible treatment moderators such as problem severity, ethnicity, number of treatment sessions, mode of treatment delivery and methodological issues. Some of these variables are important from a descriptive standpoint, while others are included for their potential treatment outcome implications. This distinction will be elaborated upon later in the review.
CHAPTER II
LITERATURE REVIEW

Introduction to Child Psychotherapy Research

Interest in the treatment of childhood and adolescent mental health problems has led to a series of meta-analytic reviews which have investigated not only the efficacy of treatment but also a host of factors believed to influence treatment outcomes (Casey & Berman, 1985; Weisz et al., 1987; Weisz et al., 1995). These reviews have commonly grouped internalizing and externalizing disorders together. The attempts which have been made to analyze separately the two types of presenting problems have resulted in similar effect sizes (Casey & Berman, 1985; Weisz et al., 1987; Weisz, et al., 1995). For example, Weisz et al. (1987) obtained an average effect size of .79 for externalizing problems and .88 for internalizing problems. A more recent meta-analytic review conducted by Weisz et al. (1995) resulted in an average effect size of 0.58 for externalizing and 0.44 for internalizing disorders. Although both types of problems appear to be treated with equal rates of success, very little is known about how different factors may affect outcome for the two disorders.
Comorbidity

The matter of comorbidity among children and adolescents has been the focus of recent debate and concern. Comorbidity refers to the co-occurrence of two or more disorders, and it is common among children and adolescents. Most researchers and clinicians recognize the frequent dual diagnosis of anxiety disorders and depression and of attention deficit and conduct disorders. These pairs of disorders represent clusters of symptoms which belong to the same broadband category: internalizing disorders or externalizing disorders. However, comorbid diagnoses among children and adolescents frequently include disorders from both broadband categories. For example, Weiss and Catron (1994) found a strong relation between aggressive and depressive symptomatology. In addition, Cole and Carpentieri (1990) found a strong correlation of .73 between conduct problems and depressive symptoms.

Epidemiologic studies have found that comorbidity occurs in about 50% of diagnosed children and adolescents (Anderson, Williams, McGee, & Silva, 1987). The high frequency with which dual diagnosis occurs in the general population of children and adolescents has relevance for treatment outcome research. For instance, Weisz and Weiss (1989) took issue with the high effect sizes generated from meta-analyses of child psychotherapy pointing out that most children who receive services in clinic settings are
referred for multiple presenting problems. However, the vast majority of research on child psychotherapy tends to involve the treatment of one presenting problem for which children are often recruited rather than clinic-referred (Weisz & Weiss, 1989). Therefore, Weisz and Weiss (1989) posit that very little is known about the effects of treatment for the actual population of children commonly receiving services.

The issue of comorbidity or multiple presenting problems will be empirically addressed in this investigation. Studies of interventions for children presenting with mixed problems will be examined both descriptively and with respect to treatment outcome. Comorbid studies will be analyzed separately and compared to those studies which were designed to treat a single internalizing presenting problem.

Gender Issues

Recent epidemiological studies of internalizing disorders have produced an interesting and complex finding regarding prevalence rates for boys and girls. Girls seem to suffer distress from internalizing disorders at rates much higher than those for boys, particularly during the adolescent years. For example, in the study of epidemiological rates for anxiety and phobic disorders, Anderson (1994) reported that girls tended to present more frequently with anxious symptoms, especially among older
children. Phobic symptoms and somatic concerns such as recurrent abdominal pain have also been found to occur more frequently in girls (Abe & Matsui, 1981; Feldman, Hodgson, Corber, 1985). A similar pattern exists for rates of depression, with more females being identified among adolescents (Kaplan, Hong, & Weinhold, 1984; Reynolds, 1985).

This evidence suggests a fairly consistent trend for girls to experience more distress from internalizing symptoms than boys, and this trend seems to become more observable in adolescence. The influence of gender will be examined in this review. Descriptive information will be provided to determine if outcome research follows general epidemiological patterns (that is, are more females than males treated in studies of internalizing disorders?). Furthermore, the effect sizes will be compared for the studies treating primarily males and those treating primarily females. Past meta-analytic reviews have found that studies with a majority of girls responded better to treatment than studies which primarily included boys (Casey & Berman, 1985; Weisz et al., 1987; Weisz et al., 1995). This finding will be tested to determine if it holds true for the treatment of internalizing disorders.

Age

Longitudinal and cross-sectional studies of internalizing disorders have found that both anxiety and
depressive disorders increase with age (Ollendick & King, 1994). In addition, evidence exists that the specific type of internalizing disorder also changes with age (Ollendick & King, 1994). Anxiety disorders tend to precede mood disorders in development (Kovacs, Feinberg, Crousse-Novak, Paulauskas, & Finkelstein, 1984). Younger children are frequently diagnosed with anxiety disorders such as separation anxiety and simple phobias, but these problems typically give way to other disorders in older children and adolescents such as overanxious disorder, social phobias, and depression (Ollendick & King, 1994). Age variations for anxiety and depressive disorders will be examined in this review in terms of their prevalence and influence upon treatment outcome.

Prior meta-analytic reviews have studied different age groups to determine if differences in treatment effectiveness exist. The reviews have yielded conflicting results. Whereas one review (Weisz et al., 1987) found an effect size of 0.92 for children, and only 0.58 for adolescents, a second meta-analytic review by the same research group obtained an opposite result: an effect size of 0.48 for younger children and 0.65 for adolescents (Weisz et al., 1995). Given the conflicting findings, it is important for future research to study possible age differences in outcome.
Type of Treatment

The findings regarding the efficacy of various types of treatment have been unequivocal. Behavioral interventions consistently result in larger effect sizes than non-behavioral treatments (Casey & Berman, 1985; Weisz et al., 1987; Weisz et al., 1995). Some concern was raised by Casey and Berman (1985) that behavioral treatments had an advantage over non-behavioral ones in that the outcome measures were frequently quite similar to the activities of therapy. Upon removing these measures, the superior treatment effects for non-behavioral interventions disappeared (Casey & Berman, 1985). Weisz et al. (1987; 1995) conducted a similar analysis in their reviews, but they only removed those outcome measures which were unnecessarily similar to the activities of treatment. Their results suggested that behavioral interventions were still superior to the non-behavioral treatments (Weisz et al., 1987; Weisz et al., 1995).

The finding that behavioral interventions are more successful than non-behavioral treatments hold constant even when type of presenting problem was considered (internalizing or externalizing) (Weisz et al., 1987; Weisz et al., 1995). In their discussion of treatment types, Weisz et al. (1995) noted that externalizing (undercontrolled) problems such as conduct disorder and attention deficit are often treated with behavioral
interventions; however, little descriptive information was provided regarding the types of interventions most frequently used to treat internalizing problems. Although it is likely that research on the treatment of internalizing disorders has included many behavioral interventions, differences may occur for different types of internalizing problems. For example, phobias may be treated behaviorally more frequently than depression. This review will descriptively explore the types of treatments most commonly used for each type of internalizing disorder; it will also examine the effect sizes for the different treatment types.

Type of Therapist

Prior meta-analytic reviews have failed to find a main effect for type of therapist or therapist training (Weisz et al., 1987; Weisz et al., 1995). It appears that therapy outcome is not related to how many years of experience a therapist has; furthermore, paraprofessionals seem to be as effective at treating children as professionals (Weisz et al., 1987; Weisz et al., 1995). However, an interesting interaction was found when the type of therapist was examined separately for internalizing and externalizing problems. It seems that externalizing problems are treated equally well by both professionals and paraprofessionals; however, professionals treat internalizing problems with significantly higher levels of success than paraprofessionals and graduate students (Weisz et al., 1987;
Weisz et al., 1995). The reported effect sizes were .86 for professionals, .14 for paraprofessionals, and .56 for graduate student therapists (Weisz et al., 1995).

Weisz et al. (1995) offered the following explanation for the interaction between presenting problem and type of therapist: "...the kinds of behavior management interventions often used with undercontrolled problems tend to be clear cut enough to be taught efficiently to parents and teachers through a focused training program but the interventions needed for the more subtle and less overt problems that tend to fall within the overcontrolled category do indeed require substantial professional training" (Weisz et al., 1996, p.462). This review will investigate the types of therapists and treatments used to treat internalizing disorders to test this finding. Variables such as therapist training and experience will be tested as possible moderators of treatment outcome.

Source of Outcome Measure

The sources used to evaluate the effectiveness of child treatment have typically included teachers, parents, clinicians, independent observers, peers, and self-reports. Past meta-analytic reviews have found some differences regarding the effects obtained from these different sources. For example, Casey and Berman (1985) found significantly higher effect sizes for measures obtained from parents, therapists, and independent observers than those obtained
from the child or the child's teacher. Weisz et al. (1987) found higher effect sizes for observers than for all other sources. This review will also compare the effect sizes for the different sources of outcome measures; however particular attention will be given to the reports given by parents, teachers, and the children themselves.

Given the extensive amount of time teachers spend with children, they are often consulted in evaluations of treatment effectiveness. In the past, teachers ratings have been most commonly used to identify children with externalizing problems (Pelham, Gnagy, & Milich, 1992); furthermore, teachers have been fairly successful at accurately assessing externalizing symptoms (Achenbach, McConaughy, & Howell, 1987; Phares, Compas, & Howell, 1989). In addition to teachers, parents are another obvious source of information regarding the mental health of children and adolescents. The referral and assessment of childhood disorders has typically relied heavily upon information obtained from parents. However, similar to teachers, parents are more successful at assessing externalizing symptoms (Phares et al., 1989).

In various clinical studies, concordance among parent, teacher, and child reports of the presence of internalizing problems, has been surprisingly low (Epkins, 1993; Kazdin, Esvelt-Dawson, Unis, Rancurello, 1983; Mokros, Poznanski, Grossman, & Freeman, 1987; Phares et al., 1989; Sacco &
Graves, 1985; Stavrakaki, Vargo, Roberts, & Boodoosingh, 1987). For example, teacher ratings have resulted in only weak correlations with child self-reports of depressive and anxious symptomatology (Epkins, 1993; Phares et al., 1989; Sacco & Graves, 1985). This finding has been consistent in both clinically-referred samples as well as elementary school samples. In a study conducted by Epkins (1993), samples of clinically-referred children and elementary school children both rated their internalizing symptoms as more distressing than teachers. Other studies simply found no relationship between the teacher and child reports (Phares et al., 1989; Sacco & Graves, 1985).

The results of these studies provide little support for the use of teacher reports for identifying children in distress due to internalizing problems or for assessing changes in distress levels following treatment. There is some evidence that teachers are able to gage the general severity of disturbances, particularly for referred children even if they cannot identify the exact symptoms present (Epkins, 1993). This is likely due to the more subtle symptoms of internalizing disorders which are not directly observable through normal classroom interactions and may be viewed as less problematic by teachers than acting-out types of behaviors. Therefore, reliance on teacher reports in assessing improvements following treatment for internalizing disorders may be problematic. This investigation will
examine the issue of teacher reports for internalizing symptomatology by comparing teacher-provided outcome data to other sources of outcome data.

In addition to teacher reports, correspondence between parental and child reports has not been high (Kazdin et al., 1983; Mokros et al., 1987; Phares et al., 1989; Stavrakaki et al., 1987). In a study conducted by Mokros et al. (1987), a sample of both referred and non-referred children was used. The correlations between child and parent reports were low for both groups; however, the direction of the discrepancy differed for the two samples. The group of non-referred children rated their internalizing symptoms as more severe than the parental ratings; however, the clinic sample of children rated their symptoms as less severe than parents. In another study of inpatient children (Kazdin et al., 1983), a similar result was found. Children rated their depressive symptoms as less severe than parents (Kazdin et al., 1983). This finding was confirmed in another clinic sample of depressed children; however, it did not hold true for anxious children (Stavrakaki et al., 1987). The anxious children tended to rate their internalizing symptoms as more severe than parents.

These findings imply that while concordance among children and parents is strikingly low; the discrepancy may be moderated by the type of sample. Referred children either exhibit a tendency to minimize their symptoms, or the
discrepancy may be due to parents who exaggerate their
childrens' symptoms. Conversely, non-referred children seem
to either exaggerate their level of distress or are unable
to convey their distress to their parents. This interaction
will be explored; in addition, the use of parental reports
as an outcome measure will also be examined in this review.

It is clear from the research which has been conducted
comparing ratings of teachers, parents, and children that
each source brings a unique perspective to the evaluation of
child and adolescent mental health. Given the low
concordance among sources, it is particularly important that
multiple sources be considered in the assessment and
treatment of children with internalizing disorders. This
study will examine how well and consistently this has
occurred in child outcome research.

**Type of Outcome Measure**

In addition to the source of outcome measures, the type
of measure can also be a salient moderating variable. Casey
and Berman (1985) found significant differences in effect
sizes for types of outcome measures. Measures of fear and
cognitive performance generated significantly higher effect
sizes than measures of self-esteem and personality.
However, no definitive conclusions could be drawn from these
analyses since a confound existed between type of treatment
and type of outcome measure. Casey and Berman (1985) found
that the behavioral treatments which obtained higher effects
were evaluated primarily by measures of fear and cognitive performance (Casey & Berman, 1985). Therefore, it is difficult to ascertain the influence of type of outcome measure upon treatment outcome.

There are several important considerations related to internalizing disorders that need to be addressed. As noted earlier, the symptoms of internalizing disorders are often more subtle and covert than the symptoms of externalizing disorders (Reynolds, 1992). Therefore, outcome measures such as direct behavioral observations may not be as appropriate for evaluating therapeutic change. In addition, the measures used to assess progress for internalizing problems like depression and anxiety may include more stable aspects of personality such as self-esteem and level of social interaction which may be significantly more resistant to change. The fact that Casey and Berman (1985) found measures of personality and self concept to be the most resistant to change provides some preliminary evidence that a significant difference may exist. Therefore, it will be necessary to explore the types of outcome measures used to evaluate interventions for internalizing disorders both descriptively and statistically.

While the variables which have been discussed are of particular interest to this review, prior research has suggested that other factors are salient to the outcome of psychotherapy with children as well. These variables will
not be overlooked in this review, although their influence will not be as extensively explored. These additional variables are briefly discussed in the next section.

**Problem Severity**

The severity of the presenting problem is a variable which has not been explored in child psychotherapy meta-analyses (Casey & Berman, 1985; Weisz et al., 1987; Weisz et al., 1995). Therefore, it is uncertain to what extent it will influence or moderate treatment outcome. However, it may be a potent variable for the study of internalizing disorders given that these problems may not come to the attention of mental health professionals until they reach more serious levels. In addition, problems of a more serious nature represent a significant challenge for therapists given that they may be quite difficult to treat. For this reason, this variable will be examined in the current investigation.

**Ethnicity**

The ethnicity of children who are treated in child psychotherapy research has largely been ignored. Many studies of child treatment fail to disclose the ethnicity or race of the children included (Kazdin, Bass, Ayers, & Rodgers, 1990). As Kazdin (1993) noted, ethnic minority children are at higher risk for dysfunction but may have little access to mental health services. In addition, minority children are typically not well-represented in
clinic practices or in treatment research (Kazdin et al., 1990). In fact, one of the identified directions for future research is to design and evaluate interventions which are culturally sensitive to diverse populations (Kazdin, 1993). Therefore, ethnicity will be studied in this review as a potential moderating variable. However, it is uncertain to what extent it may influence the treatment of internalizing disorders.

**Number of Treatment Sessions**

The number of sessions included in an intervention has the potential to influence outcome substantially. Therefore, this variable has been explored in the meta-analytic review conducted by Casey and Berman (1985). Logic might suggest that the interventions which have more sessions will result in larger effect sizes; however, this was not the case. The review actually found a negative relationship between the number of sessions and effect size (Casey and Berman, 1985). However, it is likely that a confound existed between type of treatment and number of sessions. Brief interventions tended to be behavioral and to include outcome measures similar to the activities of therapy which typically result in larger effect sizes. It is conceivable that the treatment of internalizing disorders might require interventions which are lengthier given that the symptoms are of a more private and internally focused
nature. Therefore, the influence of number of sessions will be examined in this review.

**Mode of Treatment Delivery**

Research on child psychotherapy has primarily included individual and group therapy; however, some studies incorporate both of these modes of treatment delivery. Although the mode of treatment delivery may be a salient moderator of outcome, it has received only minimal attention in meta-analytic reviews. Weisz et al. (1987) found a larger effect size for individual than for group treatment, but a later review (Weisz et al., 1995) obtained similar effects for both treatment modalities. Mode of treatment delivery may have special significance for the treatment of internalizing disorders given that the characteristics of internalizing symptomatology are of a more personal nature. For example, children and adolescents may be more reluctant to discuss their feelings and concerns in a group setting.

Treatment modality will be evaluated in this review to determine if one mode of treatment is superior to another in the treatment of internalizing disorders.

**Methodological Issues**

While meta-analytic reviews of child psychotherapy have affirmed its effectiveness, some authors have taken issue with the methodology of the studies included in these reviews (Barnett, Docherty & Frommelt, 1991; Kazdin et al., Shirk & Russell, 1992; Smyrnios & Kirkby, 1993; Weisz, Weiss
& Donenberg, 1992). Some of the important criteria for examining methodological adequacy which have generated concern include random assignment to conditions, use of no-treatment control groups, attrition rates, use of multiple outcome measures, use of a normed outcome measure, including a generalized assessment of treatment, and collecting follow-up data.

Durlak, Wells, Cotten, & Johnson (1995) examined the child psychotherapy outcome literature in terms of these criteria. Their findings suggested that many of the studies included in meta-analytic reviews of child psychotherapy contain sound design features. Furthermore, the quality of the outcome literature has improved significantly over time (Durlak et al., 1995). The studies reviewed by Durlak et al. (1995) included interventions for treating both internalizing and externalizing problems. Therefore, although it is likely that results similar to Durlak et al. (1995) will be obtained, an analysis of methodological characteristics will be conducted for this sample of studies.

Rationale for Current Investigation

Identification Issues

Internalizing problems are significantly less likely to come to the attention of parents and school personnel for a referral (Silverman & Kearney, 1991). Children with internalizing problems may be overlooked due to the personal
nature of the symptoms and the fact that the symptoms are not easily observed (Epkins, 1991).

Evidence for the oversight of internalizing problems is reflected in a study conducted by Weisz and Weiss (1991) in which they examined the most frequently referred problems in a U.S. sample for children and adolescents. Only two of the top 20 most referable target problems were of an internalizing nature (suicidal talk and appearing withdrawn). The majority of the most frequently referred presenting problems were externalizing in nature and included such things as vandalism, stealing, fighting, disobeying orders, inflicting harm on oneself or others, and setting fires (Weisz & Weiss, 1991). In addition, many of the target problems which emerged as least referred were internalizing difficulties such as obsessive behavior, feeling unloved, an overconcern with neatness and fearing poor performance (Weisz & Weiss, 1991). Taken together, these data suggest that children with internalizing problems are less likely to receive professional help.

Goals of This Review

It is clear from the above discussion that internalizing disorders represent a common clinical problem among children and adolescents, and that the identification of these problems can be difficult. Furthermore, despite the plethora of information which can be culled from the meta-analytic reviews of child psychotherapy, very little is
known about what variables moderate treatment outcome for
children and adolescents suffering from internalizing
disorders. Past reviews have hinted at possible treatment
moderators for internalizing disorders (e.g. therapist
experience, gender etc.); however, no comprehensive review
has been conducted.

The goal of this investigation is to empirically
address many of the issues pertinent to the treatment of
internalizing disorders. The general questions which will
be addressed include the following: (a) how effective are
interventions which treat internalizing disorders? (b) what
are the common characteristics of the interventions and the
children who are treated? (c) how do these characteristics
differ for the major types of internalizing disorders? and
(d) what factors are most salient as moderators of treatment
outcome? The following section will present the hypotheses
of this study.

Hypotheses

The hypotheses for this investigation will be presented
in three groups: (1) descriptive hypotheses, (2) hypotheses
related to treatment moderators, and (3) exploratory
hypotheses.

Descriptive Hypotheses

The goal of the descriptive analyses is to provide
information which summarizes the characteristics of the
children being treated and the interventions being
conducted. Several hypotheses will be made with respect to the children being treated. First, based on the demographic findings from past reviews (Abe & Matsui, 1981; Anderson, 1994; Feldman, Hodgson, Corber, 1985), it is hypothesized that more females than males will be involved in the treatment of internalizing disorders (Hypothesis 1). Second, in light of the research conducted by Ollendick and King (1994) in which age variations were found for internalizing disorders, it is hypothesized that the presenting problems of children younger than ten will primarily include symptoms of anxiety; however, older children and adolescents will be treated more frequently for somatic and depressive symptomatology (Hypothesis 2).

With regard to the interventions being tested, it is hypothesized that somatic disorders and anxiety disorders, particularly phobias, will be treated more frequently with behavioral treatments than depressive disorders (Hypothesis 3). The interventions treating depression are hypothesized to include more cognitive and non-behavioral components (Hypothesis 4).

Previous reviews have noted that behavioral treatments are used more frequently than other treatment approaches in child psychotherapy research (Casey & Berman, 1985; Kazdin et al., 1990; Weisz et al., 1987 Weisz et al., 1995). However, given that these reviews included treatment studies of both internalizing and externalizing problems, this
finding may not hold true for the exclusive study of internalizing disorders. Casey and Berman (1985) noted that treatment studies of anxiety included many behavioral interventions; however, it is uncertain if presenting problems of depression and somatization rely as heavily on behavioral components. It is hypothesized that a more representative sample of treatments will be found for this review of internalizing disorders, particularly for the treatment of depression and somatic concerns (Hypothesis 5).

Treatment Moderator Hypotheses

The next group of hypotheses being made will pertain to treatment moderators. These hypotheses will address factors previously identified like comorbidity, gender, type of treatment, type of therapist, source and type of outcome measure. The rationale for the hypotheses will be briefly outlined as well.

Given that prior meta-analytical work has found that females derive more benefits from therapy than males (Casey & Berman, 1985; Weisz et al., 1987; Weisz et al., 1995), it is believed that the effect sizes will be higher for studies containing primarily females than those containing primarily males (Hypothesis 6).

Based on the findings of Weisz et al. (1987; 1995) regarding the types of therapists, it is believed that professionals will be superior to paraprofessionals in the treatment of internalizing problems (Hypothesis 7). This
variable will be examined independently for the various
types of internalizing problems to determine if any
interactions exist.

Both the source and type of outcome measure will be
explored in this review. With regard to source of outcome
measure, it is hypothesized that the ratings of children and
teachers and of children and parents will not be
significantly correlated (Hypothesis 8). In addition, it is
hypothesized that clinically-referred children will report
less severe symptomatology than the reports from parents,
but non-referred children will report more severe distress
than parents (Hypothesis 9).

The types of outcome measures will be tested for
potential differences. In accordance with the findings of
Casey and Berman (1985), it is believed that measures of
anxiety will generate higher effect sizes than measures of
self-esteem or personality (Hypothesis 10).

While most of the hypotheses which have been made are
based on previous empirical findings, some of the analyses
which will be conducted represent new questions which have
not been addressed by previous reviews. Therefore, the
remainder of the hypotheses are considered exploratory.

Exploratory Hypotheses

Little empirical consideration has been given to the
treatment of children who are comorbid. Therefore, the
outcomes of interventions which treat multiple presenting
problems are unknown. A hypothesis will be made that the studies of comorbid children will have a lower overall effect size than those treating a single presenting problem (Hypothesis 11). The rationale for this hypothesis is based on the premise that these children will simply be more difficult to treat since they are presenting with more symptoms.

As previously discussed, the meta-analyses which have been conducted commonly group presenting problems into internalizing and externalizing. Although the effect sizes have emerged as being quite similar, it is uncertain if variations exist within types of internalizing problems. Therefore, the effect sizes for types of internalizing presenting problems will be compared. It is hypothesized that effect sizes for treatments of anxiety and somatic complaints will be higher than treatments for depression and social isolation (Hypothesis 12). The rationale for this prediction is that the distress associated with depression and social isolation may reflect underlying personality traits which are more resistant to change. Symptoms of anxiety and somatic complaints appear to be more readily amenable to treatment.

Although the variables presented for the hypotheses which have been made are of particular interest for the treatment of internalizing disorders, previous reviews have found that other variables are also potential moderators of
treatment effectiveness. Therefore, these additional variables will also be explored to determine if they are indeed salient for the outcome of treatments of internalizing disorders. They include problem severity, ethnicity, number of treatment sessions, and mode of treatment delivery, as well as other methodological variables.
CHAPTER III

METHOD

Literature Search

The studies selected for inclusion in this review were obtained through four different search methods. The first search procedure consisted of a computer search of the Psyclit database using 34 key terms to identify relevant treatment studies. The second procedure involved a manual search of 15 journals which frequently publish child psychotherapy research (See Appendix A). The third procedure entailed searching all references from identified studies as well as the references from previous meta-analytical reviews. Finally, a computer and manual search of Dissertation Abstracts was conducted. From this search, a representative sample of unpublished doctoral dissertations was obtained.

From these search procedures, studies were selected if the presenting problem of the children being treated was of an internalizing nature (i.e. anxiety, depression, social isolation, or somatic concerns). An additional search of the larger pool of studies was conducted to identify studies which treated children with mixed symptoms of which at least one was of an internalizing nature.
The final pool of studies met the following criteria: (1) the treated children or adolescents had a mean age of 18 or younger; (2) there was a control group drawn from the same population as the treated group; and (3) treated children or adolescents had only internalizing problems or had internalizing problems in combination with other problems.

Many of the studies which were evaluated contained more than one treatment group. In these cases, each intervention was coded separately. The final pool of studies consisted of 155 studies yielding 178 separate interventions.

Coding Procedures

Every intervention was coded on 47 variables which were divided into 7 separate sections (See Appendix B). The sections coded the following characteristics: (1) the study (i.e. year of publication, type of intervention); (2) design; (3) sample; (4) therapists; (5) comparison (i.e. treatment group or control group); (6) treatment; (7) outcome measures; and (8) effect size information.

Calculation of Effect Sizes

The following formula was used to calculate effect sizes:

\[ \frac{M_t - M_c}{SD\text{ pooled}} \]

The \( M_t \) represents the mean of the treatment group; \( M_c \) is equal to the mean of the control group, and \( SD\text{ pooled} \) denotes the pooled standard deviation of both groups (Hedges
& Olkin, 1985). Higher positive effect sizes denote more successful interventions. In contrast, a negative effect size indicates a stronger effect for the control group. In instances when the means and/or standard deviations are not provided, alternative procedures were used to estimate effect sizes in accordance with Wolf (1986).

Additional procedures were used to (1) protect against small sample size bias; and (2) to weight effect sizes according to their respective sample sizes (Hedges & Olkin, 1985). Effect sizes for small samples were corrected to adjust for small sample bias. The weighting procedures conducted gave greater weight to those interventions which included larger samples and provided more reliable estimates of true population effects.

Analysis of Interventions

For the initial analysis, a single effect size was calculated for each separate intervention. For interventions which included more than one outcome measure, effect sizes were averaged to yield a single effect size.

Homogeneity of effect size analyses were computed for all variables believed to be potential moderators of outcome. These included 1) type of internalizing problem, 2) type of treatment, 3) type of therapist, 4) severity of presenting problem, 5) mode of treatment delivery, 6) race of children 7) age 8) gender of children being treated, 9) source of outcome data 10) type of outcome measure, and 11)
the type of adjustment being measured. In addition, homogeneity was also calculated for two important interactions: problem type by treatment type and problem type by therapist type.

The calculation of effect sizes followed Hedges and Olkin's (1985) categorical fixed effects model. This involved the calculation of a Q (goodness-of-fit) statistic for each study grouping. The $Q_{within}$ score indicates whether the effect size in each cell is homogeneous. Homogeneity indicates that the variance produced by the group of studies contained in the cell is more likely due to random error and not to systematic differences among the studies. Ideally, each variable will produce a nonsignificant $Q_{within}$. However, it is expected that only those variables which have been grouped accordingly for the between-group analysis will result in the appropriate $Q_{within}$ score. A table of critical Q values was consulted to assess the homogeneity for each study grouping.
CHAPTER IV

RESULTS

Sample Characteristics

Table 1 presents sample characteristics of the 178 interventions. The presenting problems were as follows: 27.5% involved anxiety disorders, 21.3% involved social isolation; 12.4% were somatic concerns; 7.3% phobias; and 4.5% were depression. The remaining 26.9% of studies targeted children with multiple presenting problems: 11.2% had multiple internalizing problems and 15.7% had internalizing and externalizing problems.

Of the 178 interventions, 37.1% treated children with mild symptomatology; 31.5% included children with moderate problems, and the rest of the interventions (31.5%) did not provide enough information about the presenting pathology to estimate level of severity. Descriptive results for race indicated that sixteen interventions (9%) were studies of Caucasian children; another 10 (5.6%) treated non-whites; 9 (5.1%) included mixed samples, and in the remaining 143 interventions (80.3%), the samples were of an unknown racial background. In 13.5% of the studies, children had academic difficulties; only 1.7% of the studies ruled out academic problems in their samples, and the majority of studies
Table 1. -- Sample Characteristics of Reviewed Studies

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of Problem</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>49</td>
<td>27.5</td>
</tr>
<tr>
<td>Social Isolate</td>
<td>38</td>
<td>21.3</td>
</tr>
<tr>
<td>Somatic</td>
<td>22</td>
<td>12.4</td>
</tr>
<tr>
<td>Phobia</td>
<td>13</td>
<td>7.3</td>
</tr>
<tr>
<td>Depression</td>
<td>8</td>
<td>4.5</td>
</tr>
<tr>
<td>Mixed (Internalizing)</td>
<td>20</td>
<td>11.2</td>
</tr>
<tr>
<td>Mixed (Internalizing &amp; Externalizing)</td>
<td>28</td>
<td>15.7</td>
</tr>
<tr>
<td><strong>Problem Severity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td>66</td>
<td>37.0</td>
</tr>
<tr>
<td>Moderate to Severe</td>
<td>56</td>
<td>31.5</td>
</tr>
<tr>
<td>Unknown</td>
<td>56</td>
<td>31.5</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger than seven</td>
<td>24</td>
<td>13.4</td>
</tr>
<tr>
<td>Between seven and eleven</td>
<td>82</td>
<td>46.1</td>
</tr>
<tr>
<td>Older than eleven</td>
<td>57</td>
<td>32.2</td>
</tr>
<tr>
<td>Unknown</td>
<td>15</td>
<td>8.3</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>16</td>
<td>9.0</td>
</tr>
<tr>
<td>Minority</td>
<td>10</td>
<td>5.6</td>
</tr>
<tr>
<td>Mixed</td>
<td>9</td>
<td>5.1</td>
</tr>
<tr>
<td>Unknown</td>
<td>143</td>
<td>80.3</td>
</tr>
<tr>
<td><strong>Academic Problems</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>24</td>
<td>13.5</td>
</tr>
<tr>
<td>Absent</td>
<td>3</td>
<td>1.7</td>
</tr>
<tr>
<td>Unknown</td>
<td>151</td>
<td>84.8</td>
</tr>
</tbody>
</table>
(84.8%) did not present any information regarding children's academic problems.

Gender was examined by calculating the mean percent of males contained in the samples (hypothesis 1). The mean was 47% with a standard deviation of 24 suggesting that many studies contained equal numbers of males and females. Gender was also examined for the different presenting problems. The percent of males for different problems were as follows: 35% for somatic concerns, 44% for anxiety, 49% for social isolates, 50% for depression, 53% for mixed internalizing problems, 56% for mixed internalizing and externalizing problems, and 57% for phobias. An analysis of variance of gender across problems was not significant (F=1.35; p > .05). Therefore, there was no support for the hypothesis that more females would be treated for internalizing problems than males.

The mean age of the treated children was 10.29 with a standard deviation of 3.40. Mean ages for the different presenting problems were as follows: 7.57 for social isolation, 8.20 for phobias, 9.77 for multiple presenting problems, 11.52 for anxiety, 11.84 for somatic concerns, and 13.62 for depression. An analysis of variance was conducted to determine if significant differences existed for children's ages across presenting problems (hypothesis 2). Studies of children with mixed presenting problems were dropped from this analysis. The ANOVA was significant
Post-hoc Scheffe tests conducted at the .05 level of significance revealed that children with social isolation and phobias were significantly younger than children with anxiety, depression, or somatic concerns. Otherwise, there were no significant between group differences. The findings from these analyses provide support for the hypothesis that older children will be treated more frequently for depression and somatic complaints; however, analyses did not support the hypothesis that younger children will be treated more frequently for anxiety. In this sample, younger children were most commonly treated for phobias and social isolation.

Treatment Characteristics

Descriptive characteristics of the treatments are presented in Table 2. Fifty-seven of the studies (32%) were secondary prevention. The other 68% were interventions to treat children with identified problems. Eighty of the 178 interventions (44.9%) were behavioral; 29 (16.3%) were cognitive-behavioral, and 69 (38.8%) were non-behavioral. The percentage of behavioral interventions contained in the current review is considerably lower, and the percentage of non-behavioral interventions is considerably higher than corresponding figures from two earlier child meta-analyses (Weisz et al., 1987; Weisz et al., 1995). For example, the sample of studies for Weisz et al. (1987) was 72% behavioral, 8% cognitive-behavioral and 20% non-behavioral.
Table 2.--Treatment Characteristics of Reviewed Studies

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of Treatment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral</td>
<td>80</td>
<td>44.9</td>
</tr>
<tr>
<td>Cognitive-Behavioral</td>
<td>29</td>
<td>16.3</td>
</tr>
<tr>
<td>Non-Behavioral</td>
<td>69</td>
<td>38.8</td>
</tr>
<tr>
<td><strong>Type of Therapist</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professionals</td>
<td>73</td>
<td>41.0</td>
</tr>
<tr>
<td>Professional Trainees</td>
<td>26</td>
<td>14.6</td>
</tr>
<tr>
<td>Paraprofessionals</td>
<td>33</td>
<td>18.0</td>
</tr>
<tr>
<td>Mixed</td>
<td>25</td>
<td>14.0</td>
</tr>
<tr>
<td>Unknown</td>
<td>22</td>
<td>12.4</td>
</tr>
<tr>
<td><strong>Treatment Modality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual</td>
<td>58</td>
<td>32.4</td>
</tr>
<tr>
<td>Group</td>
<td>104</td>
<td>58.3</td>
</tr>
<tr>
<td>Mixed</td>
<td>13</td>
<td>7.4</td>
</tr>
<tr>
<td>Unknown</td>
<td>3</td>
<td>2.0</td>
</tr>
<tr>
<td><strong>Number of Sessions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 or less</td>
<td>124</td>
<td>69.6</td>
</tr>
<tr>
<td>Between 11 and 20</td>
<td>37</td>
<td>20.8</td>
</tr>
<tr>
<td>Between 21 and 50</td>
<td>14</td>
<td>7.9</td>
</tr>
<tr>
<td>More than 50</td>
<td>3</td>
<td>1.7</td>
</tr>
</tbody>
</table>
The second meta-analysis (Weisz et al., 1995) contained 71% behavioral interventions, 17% cognitive-behavioral and 12% non-behavioral interventions. Therefore, the hypothesis that treatments for internalizing disorders would contain more non-behavioral treatment approaches was supported (hypothesis 5).

Chi-square analyses were conducted to test the hypothesis that a higher frequency of interventions treating somatic problems, anxiety disorders and phobias would be behavioral (hypothesis 3), and that social isolation and depression would contain more non-behavioral and cognitive-behavioral components (hypothesis 4). The results from this analysis indicated that no significant differences existed ($\chi^2=.06151, p > .05$). The percentages of behavioral, cognitive-behavioral, and non-behavioral interventions appear to be the same for the different presenting problems; therefore, hypotheses 3 and 4 were not supported.

Therapists included mental health professionals (41%), professional trainees (14.6%), and paraprofessionals such as parents or teachers (18%). Another 14% combined therapists from these categories, and the remaining 12.4% utilized therapists with unknown training. Of the 178 interventions, 34.4% were individual treatment, 58.3% were conducted in groups, and 7.4% included both an individual and a group component.
The mean number of treatment sessions was 11.43 with a standard deviation of 14.12. Both the median and the mode for number of sessions was eight. There was a non-significant correlation between the number of sessions and average effect size of -0.14. Many of the interventions were brief. For example, 69.6% of the interventions involved 10 or fewer sessions. Another 20.8% of the studies lasted from 11 to 20 sessions, and 7.9% involved between 21 and 50 sessions. Only 1.7% of the treatments were longer than 50 sessions. The possibility that a quadratic relationship might exist between number of sessions and outcome was investigated (i.e. very brief interventions and very lengthy interventions might have higher effect sizes than interventions with medium numbers of sessions). However, the analyses found no evidence for such a trend.

**Methodological Features**

Table 3 contains information regarding several methodological features of the studies. Random assignment to conditions was present in 79% of the interventions. Attention placebo control groups were used in 27.9% of the studies. Attrition rates for 88.7% of the studies were less than 10%. Follow-up data was collected in 32% of the studies. At least one normed outcome measure was present in 30.2% of studies, and 61.2% included a measure designed to assess whether treatment had an impact across behaviors or settings. Most of the studies (65.2%) used more than one
Table 3.—Methodological Characteristics of Reviewed Studies

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of Design</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonequivalent Control Group</td>
<td>24</td>
<td>13.5</td>
</tr>
<tr>
<td>Randomized True Experiment</td>
<td>140</td>
<td>78.7</td>
</tr>
<tr>
<td>Other</td>
<td>14</td>
<td>7.9</td>
</tr>
<tr>
<td><strong>Type of Control Group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Treatment</td>
<td>100</td>
<td>55.9</td>
</tr>
<tr>
<td>Wait List</td>
<td>28</td>
<td>16.2</td>
</tr>
<tr>
<td>Attention Placebo</td>
<td>50</td>
<td>27.9</td>
</tr>
<tr>
<td><strong>Type of Outcome Measure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-normed Rating Scale</td>
<td>60</td>
<td>33.9</td>
</tr>
<tr>
<td>Normed Rating Scale</td>
<td>40</td>
<td>22.5</td>
</tr>
<tr>
<td>Independent Behavioral Observation</td>
<td>30</td>
<td>16.9</td>
</tr>
<tr>
<td>Non-academic Performance Measure</td>
<td>20</td>
<td>11.5</td>
</tr>
<tr>
<td>Achievement Test or IQ Measure</td>
<td>14</td>
<td>7.9</td>
</tr>
<tr>
<td>Peer Sociometric</td>
<td>14</td>
<td>7.3</td>
</tr>
<tr>
<td><strong>Source of Outcome Measure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject Self-report</td>
<td>59</td>
<td>33.0</td>
</tr>
<tr>
<td>Teachers</td>
<td>29</td>
<td>16.2</td>
</tr>
<tr>
<td>Independent Observers</td>
<td>24</td>
<td>13.8</td>
</tr>
<tr>
<td>Subjective Performance Measures</td>
<td>20</td>
<td>11.2</td>
</tr>
<tr>
<td>Parents</td>
<td>14</td>
<td>7.9</td>
</tr>
<tr>
<td>Peers</td>
<td>13</td>
<td>7.3</td>
</tr>
<tr>
<td>Other</td>
<td>13</td>
<td>7.3</td>
</tr>
<tr>
<td>Mixed</td>
<td>6</td>
<td>3.3</td>
</tr>
<tr>
<td><strong>Type of Adjustment Measured</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral</td>
<td>70</td>
<td>39.7</td>
</tr>
<tr>
<td>Personality</td>
<td>49</td>
<td>27.4</td>
</tr>
<tr>
<td>Academic Performance</td>
<td>19</td>
<td>10.7</td>
</tr>
<tr>
<td>Physiological</td>
<td>17</td>
<td>9.5</td>
</tr>
<tr>
<td>Peer Sociometric</td>
<td>15</td>
<td>8.7</td>
</tr>
<tr>
<td>Cognitive</td>
<td>4</td>
<td>2.0</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>2.0</td>
</tr>
<tr>
<td><strong>Dimension of Adjustment Measured</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Adjustment/Social Skills</td>
<td>56</td>
<td>31.3</td>
</tr>
<tr>
<td>Fear/Anxiety</td>
<td>46</td>
<td>25.9</td>
</tr>
<tr>
<td>Personality</td>
<td>22</td>
<td>12.6</td>
</tr>
<tr>
<td>Achievement</td>
<td>17</td>
<td>9.3</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>15</td>
<td>8.6</td>
</tr>
<tr>
<td>Physiology</td>
<td>13</td>
<td>7.2</td>
</tr>
<tr>
<td>Cognitive Skills</td>
<td>9</td>
<td>5.1</td>
</tr>
</tbody>
</table>
outcome measure, and the mean number of outcome measures per study was 2.8 with a standard deviation of 1.6.

Of the 303 outcome comparisons, 33% were subject self-report, 16.2% were obtained from teachers, 13.9% were obtained from independent observers, 7.9% from parents, and 7.3% from peers. Another 11.2% were experimenter-constructed performance measures, and the remaining 10.5% were other or mixed.

The hypothesis that a low concordance rate would exist between child report and parent report as well as between child report and teacher report was explored (hypothesis 8). Studies for which child and parent information were available were selected, and a correlation was conducted. The resulting correlation between child and parent report was 0.49. Since this figure was based on only seven studies, it was non-significant. For these studies, the mean effect size for children was 0.33, and it was 0.20 for parents. A similar analysis was conducted to examine the concordance between child and teacher report, and the correlation was 0.05 and was also non-significant (n=27). For this group of studies, the mean effect size for child reports was 0.61, but it was only 0.30 for teacher reports. In both cases, it appears that children report more benefits from treatment then either parents or teachers, and the discrepancy is larger for studies examining teacher verses child reports. Taken together, these analyses supported
hypothesis eight that the concordance between child and teacher reports would be low; however, the analyses involving child and parent data only contained seven studies.

The hypothesis that clinically-referred children would report less severe symptomatology than the reports of parents could not be explored with this sample of studies (hypothesis 9). Only seven studies included a child report and parent report, and of these seven, none were samples of clinically-referred children.

The types of outcome measures fell into the following categories: non-normed or experimenter constructed instruments (33.9%), normed rating scales (22.4%), independent behavioral observations (16.8%), non-academic performance measures (11.5%), achievement tests (7.9%), and peer sociometric ratings (7.2%). The outcome measures assessed adjustment in five outcome domains: behavioral (39.6%), personality (27.3%), academic performance (10.6%), sociometric status (8.6%), and physiological functioning (9.4%). The other 3.6% did not fit into the above categories. The dimensions which were assessed included the following: social adjustment (30.8%), fear and/or anxiety (26.5%), personality (12.6%), achievement (9.6%), self-esteem (8.5%), physiology (7.1%), and cognitive skills (4.9%).
Homogeneity of Effect Sizes

To examine significance of each mean effect size, 95% confidence intervals were constructed around the mean. If the range specified by the confidence interval does not contain zero, than the mean effect size differs significantly from zero. The confidence intervals are also used to determine if cells differ significantly from each other. When the confidence intervals of two cells do not overlap, the effect sizes differ significantly at the .05 level.

When testing for homogeneity, a minimal $Q_{\text{within}}$ score is desirable because it indicates that differences within groups are most likely due to random sampling error. The size of the $Q_{\text{within}}$ statistic will fluctuate according to the number of studies within a cell. Therefore, cells with larger numbers of studies can have larger $Q_{\text{within}}$ scores and still be considered homogeneous. The final statistic of interest is the Fail safe $N$. This statistic is a reflection of the reliability of the findings for each mean effect. Fail safe $N$'s indicate the number of studies with zero effect sizes which would be necessary to reduce the obtained mean effect size to nonsignificance. Therefore, cells which have higher Fail safe $N$'s have more reliable findings than those cells which have lower Fail safe $N$'s.

In instances when homogeneity was not obtained for a particular cell, the possibility of outliers obscuring the
results was considered. In accordance with Hedge's and Olkin's model, two types of outliers are possible. These include interventions which have unusually high or low effect sizes, and interventions with unusually large sample sizes. Both types have the potential of distorting the average effect size and of inflating the Q statistic such that the cell cannot be considered homogeneous. By removing outliers from the analyses, homogeneity was obtained for some cells which were originally not homogeneous; however, no more than 5% of the studies for a particular cell were removed. Therefore, if after removing 5% of the studies as potential outliers, the cell still was still not homogeneous, the outliers were added back into the analysis, and the cell was not considered to be homogeneous.

**Treatment Outcomes**

A weighted mean effect size of 0.42 was obtained for all 178 interventions. The 95% confidence intervals were 0.37 to 0.47. The mean effect size of 0.42 indicates that the average child in a treated group was better off than 66% of the children in control groups. However, the mean effect size of 0.42 was not homogeneous suggesting the need to subdivide groups to achieve homogeneity.

Thirteen variables were investigated as potential moderators of treatment outcome. Eight of these variables were related to the hypotheses of this investigation (type of problem, age, gender, type of treatment, type of
therapist, source of outcome measure, type of outcome measure, and dimension of adjustment). The remaining five were explored in an ad hoc fashion due to their potential to influence treatment outcome (type of adjustment, problem severity, mode of treatment delivery, ethnicity, and number of treatment sessions). In addition, two interactions hypothesized to be potentially relevant were also investigated to determine their significance upon outcome (problem type by treatment type and problem type by therapist type). Results from these analyses are presented below.

**Homogeneity Results**

Table 4 summarizes the results of homogeneity analyses and presents mean effect sizes, confidence intervals and Fail safe N's for different variables. Unknown categories were dropped from tests for homogeneity and significance (e.g., race, problem severity, etc.). Homogeneity was obtained for most of the subcategories for many variables; however, it was only obtained for all categories for two variables (race and problem severity). When homogeneity is obtained, it suggests that studies have been grouped appropriately for a between-group analysis. When homogeneity is not obtained, interpretations of any group differences must be made more cautiously. In Table 4, categories which were homogeneous are indicated by an asterisk.
Table 4.—Findings from Homogeneity Analyses for Clinical Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>$Q_{within}$</th>
<th>Mean</th>
<th>95% Confidence</th>
<th>Failsafe N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Problem Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Isolate</td>
<td>38</td>
<td>93.31</td>
<td>.35</td>
<td>.24</td>
<td>.47</td>
</tr>
<tr>
<td>Phobia*</td>
<td>13</td>
<td>12.27</td>
<td>.88</td>
<td>.42</td>
<td>.61</td>
</tr>
<tr>
<td>Anxiety*</td>
<td>47</td>
<td>62.11</td>
<td>.51</td>
<td>.35</td>
<td>.47</td>
</tr>
<tr>
<td>Somatic*</td>
<td>21</td>
<td>24.60</td>
<td>.75</td>
<td>.50</td>
<td>.61</td>
</tr>
<tr>
<td>Depression*</td>
<td>8</td>
<td>8.71</td>
<td>.20</td>
<td>.50</td>
<td>.61</td>
</tr>
<tr>
<td>Multiple (internal)</td>
<td>20</td>
<td>52.56</td>
<td>.28</td>
<td>.16</td>
<td>.39</td>
</tr>
<tr>
<td>Multiple (internal and external)*</td>
<td>28</td>
<td>30.38</td>
<td>.20</td>
<td>.09</td>
<td>.32</td>
</tr>
<tr>
<td>Therapist Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional</td>
<td>73</td>
<td>160.59</td>
<td>.48</td>
<td>.40</td>
<td>.56</td>
</tr>
<tr>
<td>Professional Trainee*</td>
<td>25</td>
<td>36.81</td>
<td>.54</td>
<td>.40</td>
<td>.56</td>
</tr>
<tr>
<td>Paraprofessional*</td>
<td>32</td>
<td>51.99</td>
<td>.30</td>
<td>.21</td>
<td>.39</td>
</tr>
<tr>
<td>Mixed*</td>
<td>24</td>
<td>36.34</td>
<td>.53</td>
<td>.41</td>
<td>.65</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Under 7</td>
<td>24</td>
<td>63.65</td>
<td>.59</td>
<td>.45</td>
<td>.73</td>
</tr>
<tr>
<td>Ages 7 to 11*</td>
<td>82</td>
<td>110.69</td>
<td>.35</td>
<td>.28</td>
<td>.42</td>
</tr>
<tr>
<td>Over Age 11*</td>
<td>57</td>
<td>72.59</td>
<td>.54</td>
<td>.45</td>
<td>.63</td>
</tr>
<tr>
<td>Treatment Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral</td>
<td>89</td>
<td>217.20</td>
<td>.52</td>
<td>.45</td>
<td>.60</td>
</tr>
<tr>
<td>Cognitive-Behavioral*</td>
<td>33</td>
<td>41.80</td>
<td>.57</td>
<td>.45</td>
<td>.69</td>
</tr>
<tr>
<td>Non-Behavioral*</td>
<td>57</td>
<td>69.84</td>
<td>.27</td>
<td>.20</td>
<td>.35</td>
</tr>
</tbody>
</table>

* denotes homogeneous cell grouping
Table 4.--continued

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Q within</th>
<th>Mean</th>
<th>95% Confidence</th>
<th>Failsafe N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Dimension of Adjustment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety/Fear</td>
<td>72</td>
<td>165.09</td>
<td>.51</td>
<td>.43</td>
<td>.59</td>
</tr>
<tr>
<td>Cognitive Skills*</td>
<td>14</td>
<td>11.74</td>
<td>.39</td>
<td>.19</td>
<td>.59</td>
</tr>
<tr>
<td>Social Adjustment</td>
<td>87</td>
<td>161.55</td>
<td>.25</td>
<td>.18</td>
<td>.31</td>
</tr>
<tr>
<td>Achievement*</td>
<td>26</td>
<td>44.28</td>
<td>.23</td>
<td>.10</td>
<td>.36</td>
</tr>
<tr>
<td>Personality</td>
<td>35</td>
<td>68.51</td>
<td>.48</td>
<td>.38</td>
<td>.58</td>
</tr>
<tr>
<td>Self-esteem*</td>
<td>24</td>
<td>25.86</td>
<td>.13</td>
<td>.00</td>
<td>.25</td>
</tr>
<tr>
<td>Physiology*</td>
<td>20</td>
<td>35.90</td>
<td>.46</td>
<td>.32</td>
<td>.60</td>
</tr>
<tr>
<td>Treatment Modality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual</td>
<td>56</td>
<td>166.05</td>
<td>.46</td>
<td>.37</td>
<td>.55</td>
</tr>
<tr>
<td>Group</td>
<td>105</td>
<td>203.98</td>
<td>.42</td>
<td>.36</td>
<td>.48</td>
</tr>
<tr>
<td>Mixed*</td>
<td>12</td>
<td>14.19</td>
<td>.43</td>
<td>.25</td>
<td>.60</td>
</tr>
<tr>
<td>Problem Severity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mild*</td>
<td>65</td>
<td>93.54</td>
<td>.55</td>
<td>.47</td>
<td>.63</td>
</tr>
<tr>
<td>Moderate to Severe*</td>
<td>53</td>
<td>78.58</td>
<td>.47</td>
<td>.38</td>
<td>.56</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More Males*</td>
<td>23</td>
<td>22.17</td>
<td>.48</td>
<td>.33</td>
<td>.64</td>
</tr>
<tr>
<td>More Females*</td>
<td>20</td>
<td>32.05</td>
<td>.66</td>
<td>.49</td>
<td>.82</td>
</tr>
<tr>
<td>Equal Numbers</td>
<td>67</td>
<td>153.41</td>
<td>.53</td>
<td>.45</td>
<td>.61</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mostly Caucasian*</td>
<td>15</td>
<td>25.60</td>
<td>.34</td>
<td>.21</td>
<td>.47</td>
</tr>
<tr>
<td>Mostly Minority*</td>
<td>10</td>
<td>7.86</td>
<td>.40</td>
<td>.20</td>
<td>.61</td>
</tr>
<tr>
<td>Mixed*</td>
<td>9</td>
<td>7.63</td>
<td>.62</td>
<td>.40</td>
<td>.84</td>
</tr>
</tbody>
</table>

* denotes homogeneous cell grouping
Table 4.--continued

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Q&lt;sub&gt;within&lt;/sub&gt;</th>
<th>Mean</th>
<th>95% Confidence</th>
<th>Failsafe N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Source of Outcome Measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent observers</td>
<td>42</td>
<td>113.93</td>
<td>.37</td>
<td>.26</td>
<td>.48</td>
</tr>
<tr>
<td>Parents</td>
<td>24</td>
<td>58.61</td>
<td>.45</td>
<td>.30</td>
<td>.59</td>
</tr>
<tr>
<td>Teachers*</td>
<td>47</td>
<td>60.33</td>
<td>.25</td>
<td>.16</td>
<td>.33</td>
</tr>
<tr>
<td>Peers*</td>
<td>22</td>
<td>36.39</td>
<td>.23</td>
<td>.11</td>
<td>.36</td>
</tr>
<tr>
<td>Self-report</td>
<td>100</td>
<td>211.60</td>
<td>.35</td>
<td>.29</td>
<td>.41</td>
</tr>
<tr>
<td>Performance Measure*</td>
<td>31</td>
<td>50.75</td>
<td>.42</td>
<td>.30</td>
<td>.54</td>
</tr>
<tr>
<td>Type of Adjustment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral</td>
<td>97</td>
<td>243.22</td>
<td>.39</td>
<td>.32</td>
<td>.46</td>
</tr>
<tr>
<td>Personality</td>
<td>67</td>
<td>137.18</td>
<td>.34</td>
<td>.27</td>
<td>.41</td>
</tr>
<tr>
<td>Academic Performance*</td>
<td>25</td>
<td>33.19</td>
<td>.24</td>
<td>.13</td>
<td>.35</td>
</tr>
<tr>
<td>Sociometric Status*</td>
<td>21</td>
<td>36.18</td>
<td>.23</td>
<td>.11</td>
<td>.36</td>
</tr>
<tr>
<td>Cognitive*</td>
<td>5</td>
<td>5.30</td>
<td>.38</td>
<td>.05</td>
<td>.71</td>
</tr>
<tr>
<td>Physiology*</td>
<td>22</td>
<td>32.83</td>
<td>.78</td>
<td>.62</td>
<td>.94</td>
</tr>
<tr>
<td>Type of Outcome Measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavior Observation</td>
<td>51</td>
<td>147.04</td>
<td>.37</td>
<td>.27</td>
<td>.47</td>
</tr>
<tr>
<td>Peer Sociometric*</td>
<td>22</td>
<td>36.39</td>
<td>.23</td>
<td>.11</td>
<td>.36</td>
</tr>
<tr>
<td>Normed Rating Scale</td>
<td>68</td>
<td>151.67</td>
<td>.33</td>
<td>.26</td>
<td>.40</td>
</tr>
<tr>
<td>Non-normed Scale*</td>
<td>103</td>
<td>185.92</td>
<td>.35</td>
<td>.29</td>
<td>.41</td>
</tr>
<tr>
<td>Achievement Test*</td>
<td>23</td>
<td>34.16</td>
<td>.26</td>
<td>.13</td>
<td>.38</td>
</tr>
<tr>
<td>Other Performance Measure</td>
<td>35</td>
<td>79.13</td>
<td>.68</td>
<td>.47</td>
<td>.80</td>
</tr>
</tbody>
</table>

* denotes homogeneous cell grouping
For type of problem, homogeneity was obtained for all categories except for social isolation and multiple internalizing problems. For gender, homogeneity was reached for studies containing mostly males and studies containing mostly females; however, the group of studies containing approximately equal numbers of males and females was not homogenous. Two of the three age groups were homogeneous (ages 7-11 and over 11).

Two of the three groups of treatments were homogeneous (Cognitive-behavioral and non-behavioral). Therapist groups were homogeneous except for professionals. Only one of the three treatment modalities emerged as homogeneous (combined). For number of sessions, all groups were homogeneous except for the group of studies with less than 10 sessions. For source of outcome measure, three of the six categories were homogeneous (peers, teachers, and subjective performance measures). Homogeneity was obtained for only two of the six types of outcome measures (peer sociometric and achievement tests). For type of adjustment measured, all groups were homogeneous except for behavioral and personality changes. For dimensions of adjustment, four of the seven groups were homogeneous (cognitive skills, achievement, self-esteem, and physiology). None of the methods variables yielded homogeneous groupings; however, one variable (type of control group) resulted in one cell which was homogeneous (attention placebo).
For the interactions, most groups were homogeneous. For the interaction between type of problem and type of treatment, all groups were homogeneous except for the behavioral intervention groups. For the other interaction, type of problem and type of therapist, all groups were homogeneous with the exception of professionals treating social isolation.

Clinical Variables

Table 4 contains information regarding the mean effect sizes and confidence intervals for all clinical variables. The findings for the variables which resulted in significant between-groups differences are discussed below.

Type of Problem

The specific type of internalizing problem was found to be a significant moderator of treatment outcome. The highest effect size was obtained for the treatment of phobias (0.88) followed by 0.76 for depression, 0.75 for somatic problems, 0.51 for anxiety, 0.35 for social isolation, 0.28 for multiple internalizing problems, and 0.20 for mixed internalizing and externalizing problems. Inspection of confidence intervals indicated that treatment for phobias was significantly more effective than treatment of anxiety, social isolation, and both types of multiple problems. Studies treating depression and somatic complaints were significantly better than those treating social isolation and multiple problems. Lastly, the
treatment of anxiety resulted in a significantly higher effect size than treatments of multiple problems.

Two hypotheses were made with respect to outcomes for presenting problems. As predicted, studies treating children with multiple problems yielded lower effect sizes (hypothesis 11). The second hypothesis predicted that treatments for anxiety and somatic complaints would be more successful than treatments for depression and social isolation (hypothesis 12). This hypothesis was only partially supported. Treatments for somatic complaints were more successful than treatments for social isolation, but they were not significantly more successful than treatments for anxiety. Furthermore, contrary to the hypotheses, treatments for depression emerged as being better than treatments for social isolation as well.

Age

To test the effect of age, three age groups were established which roughly correspond to Piagetian cognitive-developmental levels. The first group (children less than seven years of age) yielded an average effect size of 0.59. Studies of children between the ages of seven and eleven had an average effect size of 0.35. The third group of children (over the age of eleven) had an average effect size of 0.54. The studies of children under the age of seven and over the age of eleven were both significantly higher than studies of
children between the ages of seven and eleven; however, they did not differ significantly from each other.

Type of Treatment

Type of treatment also emerged as a moderator of treatment outcome. The mean effect size for cognitive-behavioral treatments was 0.57. It was 0.52 for behavioral and 0.27 for non-behavioral. Both cognitive-behavioral and non-behavioral treatments were significantly more successful than nonbehavioral ones; however, they did not differ from each other.

Type of Therapist

Significant differences were also found for the types of therapists used in these interventions. The effect sizes were as follows: 0.54 for professional trainees, 0.53 for studies including therapists with mixed training, 0.48 for professionals, and 0.30 for paraprofessionals. As hypothesized, professionals, and professional trainees emerged as being significantly more successful than paraprofessionals (hypothesis 7). In addition, therapists with mixed training were also more successful than paraprofessionals. No other significant differences existed among the other therapists.

Dimension of Adjustment

Outcomes also differed significantly based on the dimension of adjustment measured. The effect sizes for the groups were as follows: 0.51 for fear and/or anxiety, 0.48
for personality, 0.46 for physiology, 0.39 for cognitive skills, 0.25 for social adjustment, 0.23 for achievement, and 0.13 for self-esteem. Examination of the confidence intervals revealed that measures of fear/anxiety and measures of personality yielded significantly higher effect sizes than measures of social adjustment, achievement, and measures of self-esteem. In addition, measures of physiology had significantly higher effect sizes than measures of social adjustment and self-esteem. Based on previous meta-analytic work, it was hypothesized that measures of fear and anxiety would result in higher effect sizes than measures of personality and self-esteem (hypothesis 10). This hypothesis was only partially supported by these results. Measures of fear/anxiety did result in higher effect sizes than measures of self-esteem; however they were not significantly higher than measures of personality.

The remaining eight clinical variables did not emerge as significant moderators of treatment outcome. Therefore, the hypothesis that interventions treating mostly females would be more successful than interventions treating mostly males was not supported (hypothesis 6). Results from these additional analyses are also presented in Table 4 along with results for other tested variables which were unrelated to the hypotheses of this investigation.
Interactions

Because treatment type has emerged as an influential variable in past meta-analytic research, the interaction of this variable and type of problem was examined. However, because the cell sizes for several presenting problems were too small (i.e., less than five studies for at least one study grouping) the results from this interaction can only be presented for two types of presenting problems: social isolation and anxiety. Table 5 presents the findings from these analyses. The only significant between-group difference was found for the treatment of social isolation where behavioral interventions emerged as significantly better than non-behavioral interventions but not significantly different than Cognitive-behavioral interventions.

The interaction between type of therapist and type of problem was also explored. Again, the cell sizes were only large enough for two of the problem types (social isolation and anxiety). There were no significant differences among the study groupings.

Methodological Variables

Homogeneity analyses were conducted on seven dichotomous methodological variables to examine the potential influence of methodology upon treatment outcome. These variables measured the presence of random assignment to conditions, attrition rates less than 10%, use of an
### Table 5.--Findings from Homogeneity Analyses for Interactions

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>$Q_{within}$</th>
<th>Mean</th>
<th>95% Confidence Low</th>
<th>95% Confidence High</th>
<th>Failsafe N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interaction Between Problem Type and Treatment Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Social Isolation:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral</td>
<td>22</td>
<td>57.62</td>
<td>.59</td>
<td>.40</td>
<td>.77</td>
<td>107.12</td>
</tr>
<tr>
<td>Cognitive-behavioral*</td>
<td>6</td>
<td>8.54</td>
<td>.28</td>
<td>.00</td>
<td>.55</td>
<td>10.51</td>
</tr>
<tr>
<td>Non-behavioral*</td>
<td>10</td>
<td>17.39</td>
<td>.20</td>
<td>.04</td>
<td>.37</td>
<td>10.19</td>
</tr>
<tr>
<td><strong>Anxiety:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral</td>
<td>24</td>
<td>52.85</td>
<td>.49</td>
<td>.35</td>
<td>.63</td>
<td>92.95</td>
</tr>
<tr>
<td>Cognitive-behavioral*</td>
<td>12</td>
<td>8.86</td>
<td>.65</td>
<td>.46</td>
<td>.83</td>
<td>65.72</td>
</tr>
<tr>
<td>Non-behavioral*</td>
<td>13</td>
<td>12.51</td>
<td>.40</td>
<td>.21</td>
<td>.59</td>
<td>38.49</td>
</tr>
</tbody>
</table>

| **Interaction Between Problem Type and Therapist Type** |
| Social Isolation: |
| Professional | 19 | 54.45 | .56 | .36 | .76 | 91.46 |
| Professional Trainee* | 5 | 4.41 | .37 | .02 | .72 | 10.72 |
| Paraprofessional* | 9 | 13.40 | .32 | .14 | .50 | 19.67 |
| Mixed | none |
| **Anxiety:** |
| Professional* | 16 | 23.41 | .41 | .25 | .58 | 50.10 |
| Professional Trainee* | 9 | 18.08 | .81 | .54 | 1.07 | 63.59 |
| Paraprofessional* | 8 | 6.74 | .33 | .09 | .58 | 18.66 |
| Mixed* | 10 | 8.46 | .59 | .39 | .78 | 48.77 |
attention placebo control group, collection of follow-up data, use of more than one outcome measure, use of at least one normed outcome measure, and use of at least one outcome measure which could assess the generalized impact of treatment.

Results from these analyses are presented in Table 6. Of the seven variables examined, three resulted in significant between group differences, but these between group differences should be interpreted with some caution since virtually all of them contained cells which were not homogeneous (see Table 5). First, studies which collected follow-up data resulted in significantly higher effect sizes than those that did not. The average effect size for studies with follow-up data was 0.55, but for studies without follow-up data, the effect size was only 0.38.

A second difference was found for studies which included a measure assessing the generalizing impact of treatment compared to those that did not include such as measure. The average effect size of 0.29 for studies including a generalized measure was significantly lower than the average effect size of 0.44 for those that did not have a generalized measure.

The third between group difference was found in the comparison of studies which included more than one outcome measure compared to those which had only one outcome measure. The average effect size of 0.30 for studies with
<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Q&lt;sub&gt;within&lt;/sub&gt;</th>
<th>Mean</th>
<th>95% Confidence</th>
<th>Failsafe N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Random Assignment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>140</td>
<td>291.41</td>
<td>.46</td>
<td>.40</td>
<td>.52</td>
</tr>
<tr>
<td>No</td>
<td>38</td>
<td>66.67</td>
<td>.33</td>
<td>.24</td>
<td>.43</td>
</tr>
<tr>
<td>Follow-up Data Available</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>57</td>
<td>105.67</td>
<td>.55</td>
<td>.45</td>
<td>.65</td>
</tr>
<tr>
<td>No</td>
<td>121</td>
<td>245.39</td>
<td>.38</td>
<td>.32</td>
<td>.44</td>
</tr>
<tr>
<td>Use of Attention Placebo Control Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes*</td>
<td>50</td>
<td>48.45</td>
<td>.41</td>
<td>.29</td>
<td>.53</td>
</tr>
<tr>
<td>No</td>
<td>128</td>
<td>324.81</td>
<td>.43</td>
<td>.38</td>
<td>.48</td>
</tr>
<tr>
<td>Attrition Rates of Less than 10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>158</td>
<td>224.50</td>
<td>.44</td>
<td>.38</td>
<td>.50</td>
</tr>
<tr>
<td>No</td>
<td>20</td>
<td>76.67</td>
<td>.45</td>
<td>.36</td>
<td>.53</td>
</tr>
<tr>
<td>Use of Multiple Outcome Measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>116</td>
<td>160.56</td>
<td>.30</td>
<td>.24</td>
<td>.36</td>
</tr>
<tr>
<td>No</td>
<td>62</td>
<td>200.27</td>
<td>.48</td>
<td>.40</td>
<td>.56</td>
</tr>
<tr>
<td>Use of a Normed Outcome Measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>54</td>
<td>125.46</td>
<td>.33</td>
<td>.26</td>
<td>.39</td>
</tr>
<tr>
<td>No</td>
<td>124</td>
<td>246.31</td>
<td>.39</td>
<td>.35</td>
<td>.44</td>
</tr>
<tr>
<td>Use of a Generalized Measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>109</td>
<td>284.59</td>
<td>.29</td>
<td>.21</td>
<td>.36</td>
</tr>
<tr>
<td>No</td>
<td>69</td>
<td>154.39</td>
<td>.44</td>
<td>.39</td>
<td>.48</td>
</tr>
</tbody>
</table>
multiple measures was significantly lower than the effect size of 0.48 for studies with only one measure.

**Multiple Regression Analyses**

Since five variables emerged as moderators of effect size, and none of these variables yielded homogeneous groups for all cells, the question of their relative influence was examined by conducting a weighted least squares multiple regression following the procedures recommended by Hedges and Olkin (1985). The five variables entered as possible predictors of effect size were presenting problem, age, type of treatment, therapist experience, and dimension of adjustment.

The average effect sizes for the three age groups indicated that the relationship between age and effect size was not linear. The correlation between age and effect size was -.11 and non-significant. Consequently, age was entered into the regression as a quadratic variable given that the relationship between age and effect size seemed to be captured more accurately as a curvilinear function.

In addition, the seven methodological variables previously evaluated were entered simultaneously at the first step of the multiple regression. This was done to examine the possibility that the methodology of the studies was a more important influence on outcome than of the clinical variables. The methodological features were entered at the first step as dichotomous variables.
indicating their presence or absence in the studies. They included random assignment to conditions, use of an attention placebo control group, attrition rate of less than 10%, availability of follow-up data, presence of a normed outcome measure, use of more than one outcome measure, and the presence of an outcome measure which assessed the generalized impact of treatment.

Table 7 presents the results of the multiple regression analyses. The block of methodological variables entered at the first step accounted for 15.28% of the variance. Type of internalizing problem entered the equation as the next best predictor of outcome ($R^2$ change = 6.45%). At the third step, experience level of the therapist was found to be the next best predictor of outcome ($R^2$ change = 3.30%). Age of the participants entered as a quadratic variable emerged as being the next most salient clinical variable for predicting outcome ($R^2$ change = 2.10%). The dimensions of adjustment entered the equation following age ($R^2$ change = 2.08%), and the final variable entered was type of treatment ($R^2$ change = 1.48%).

Therefore, the final regression model consisted of the block of methodological variables followed by the addition of five clinical variables (type of internalizing problem, therapist experience, age, dimension of adjustment, and type of treatment). Together, these variables accounted for 30.69% of the total variance.
Table 7.--Results from Regression Analyses

<table>
<thead>
<tr>
<th>Variable</th>
<th>Multiple R</th>
<th>$R^2$</th>
<th>$R^2$ change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods Variables</td>
<td>.3909</td>
<td>.1528</td>
<td>.1528**</td>
</tr>
<tr>
<td>Type of Problem</td>
<td>.4662</td>
<td>.2173</td>
<td>.0645**</td>
</tr>
<tr>
<td>Type of Therapist</td>
<td>.5002</td>
<td>.2502</td>
<td>.0330**</td>
</tr>
<tr>
<td>Age of Child</td>
<td>.5208</td>
<td>.2712</td>
<td>.0210**</td>
</tr>
<tr>
<td>Dimension of Adjustment</td>
<td>.5404</td>
<td>.2920</td>
<td>.0208*</td>
</tr>
<tr>
<td>Type of Treatment</td>
<td>.5540</td>
<td>.3069</td>
<td>.0148*</td>
</tr>
</tbody>
</table>

*p < .05  **p < .01
CHAPTER V
DISCUSSION

Efficacy of Treatments

Results from this meta-analysis confirm that treatments for internalizing disorders are effective. Some general conclusions which can be drawn are that treatment is better than no treatment, treatments are successful for a variety of internalizing problems and that treatments can be successfully implemented by different types of therapists. In addition, the high Failsafe N statistics obtained for most of the clinical variables indicate that the findings from this review are quite reliable.

Therefore, it is appropriate to conclude with a certain degree of confidence that internalizing disorders can be treated successfully. However, because several important distinctions need to be made when considering outcomes, it is useful to consider what variables are most salient for predicting outcome. The multiple regression analyses identified five clinical variables relevant for predicting outcome. The following sections will individually address each of these variables as well as discuss their implications.
Type of Problem

After controlling for methodological features, the most salient predictor for outcome was the type of internalizing problem. Somewhat contrary to hypotheses, the highest effect sizes were found for treatments of phobias (0.88), depression (0.76), and somatic problems (0.75). While the high effect sizes for phobias and somatic problems were hypothesized, the finding for depression was not expected. The successful outcomes for treatments of somatic problems and phobias might be due to the highly specific nature of their symptomatology. For example, children who have phobias may be able to function quite adequately under most circumstances. Instances when their functioning is impaired are probably highly specific, so if a treatment can remove symptoms provoked by the presence of the phobic stimulus, the child's functioning may improve rapidly to levels of normal functioning.

Conversely, problems such as anxiety and social isolation have more diffuse symptomatology (i.e., impact more areas of functioning); therefore, they might not be as amenable to treatment, particularly the brief treatments characteristic of current studies. A child who is socially isolated may have difficulty forming peer relationships. In addition, they may also have academic problems due to their lower ability to interact with teachers, and they may be withdrawn from family members as well. Therefore, treatment
will have to address more areas of distress to achieve successful outcomes.

The high effect size obtained for treatments of depression was surprising. It was originally hypothesized that treatments for depression would have lower effect sizes since the symptoms might be part of underlying personality traits that would be more resistant to change, but this was not the case. Since there were only eight studies of depression included in the review, however, the high effect size associated with treatments for depression must be interpreted cautiously. Nevertheless, the high effect sizes generated for these studies resulted in a Failsafe N of 52.56 suggesting that depression can be treated with a high degree of success.

Treatments for social isolation were less successful than other presenting problems (0.35). This finding was predicted employing a similar rationale as the one used to predict lower effect sizes for treatments of depression; that is that the symptoms associated with social isolation would be more difficult to change since they would reflect underlying personality characteristics.

Treatments for children who are comorbid (i.e., have multiple problems) are treated with much less success. The mean effect size for children with multiple internalizing problems was 0.28, while the mean for multiple internalizing and externalizing problems was 0.20. These treatments were
significantly less successful than treatments for phobias, depression; and somatic complaints. Several researchers have made the point that many children who are referred for treatment have multiple presenting problems (Kazdin, 1995; Weisz & Weiss, 1989).

If this hypothesis is true, the findings from the current review have important but discouraging implications. Research on child psychotherapy has not empirically validated highly successful treatments for children who are comorbid. There are two possibilities why the treatment of multiple problems yield lower effect sizes. First, children with multiple problems appear to have more severe pathology; therefore, it would make sense that the progress made by these children would not equal the progress made by children with single presenting problems. Second, it is possible that treatments for multiple problems require longer interventions than the ones characteristic of this sample. Most interventions (90.4%), were less than 20 sessions. While these brief interventions may be adequate for addressing single presenting problems, they may fall short of bringing about meaningful change in children with several kinds of problems.

Type of Therapist

The second clinical variable to enter the regression was type of therapist with professionals (0.48) and professional trainees (0.54) being more successful than
paraprofessionals (0.30). This result confirms the findings of Weisz et al. (1995) in their examination of therapist training for the treatment of internalizing disorders.

While this finding appears to be quite reliable, it is not clear why therapists with professional training are more adept at working with children with internalizing problems. It is also not clear if this finding is true for certain internalizing problems or for all internalizing problems. Unfortunately, due to small sample size, this review was not able to examine the interactions between therapist types and problem types adequately. More studies are needed in which the training level of the therapists is varied for the treatment of different internalizing problems. This interaction represents an important question for future research to address. In addition, future studies may wish to explore what aspects of professional training make professionals more successful in treating children with internalizing disorders.

Age

The age of children being treated emerged as being an important predictor of outcome as well. Children younger than 7 (0.59) and children older than 11 (0.54) benefit more from treatment than children between these ages (0.35). Again, it is difficult to discern why this finding occurred; however, it is possible to speculate.
The three age groups formed were intended to roughly correspond with Piagetian cognitive-developmental levels. Children in the younger age group would be considered to be at the preoperational stage of development, while the older children are considered to be at the formal operations stage of development. The middle age group (ages 7-11) would contain children at the concrete operations stage of development. It is possible that the interventions targeted to the younger children (under 7) and those targeted to older children (over age 11) were more developmentally appropriate than the interventions targeted to the middle age group.

Treatments for children between the ages of 7 and 11 may require that children utilize certain cognitive and social abilities which not all of the children have sufficiently developed. There might be more developmental variability in the middle age group which would limit some children from grasping certain concepts, particularly if the interventions require advanced problem-solving techniques or social skills training. Therefore, the outcomes might be less successful than outcomes for the other age groups.

**Dimension of Adjustment**

An important variable associated with outcome measures which emerged as a significant predictor was the type of adjustment measured. Outcome measures which assessed social adjustment (0.25), achievement (0.23), and self-esteem
(0.13) resulted in lower effect sizes than those measuring anxiety (0.51) and physiological functioning (0.46). These findings support those obtained by Casey and Berman (1985) and can be understood in terms of the constructs being measured. Social adjustment, self-esteem, and achievement reflect changes which might be said to represent the generalized impact of treatment upon functioning. Conversely, measures of anxiety and physiology are usually administered in interventions specifically designed to alter these areas of functioning. Therefore, they are measuring the specific impact of treatment as opposed to its generalized impact. Research has found that some measures of change such as self-esteem and social adjustment typically reflect smaller improvements than more specific measures (Casey & Berman, 1985).

Type of Treatment

The final clinical variable to emerge as a predictor of outcome was type of treatment. Behavioral and cognitive-behavioral treatments (0.52 and 0.57 respectively) were significantly more successful than non-behavioral treatments (0.27). Several other meta-analyses have obtained the same result (Casey & Berman, 1985; Weisz et al., 1987; Weisz et al., 1995). What is notable about this review is that the type of treatment is not as salient a variable for predicting outcome in internalizing disorders as other reviews looking at internalizing and externalizing problems
have found it to be. Other factors such as the nature of
the symptomatology and type of therapist account for more
variance in outcome.

Gender and Source of Outcome Data

Analyses failed to support two specific hypotheses made
in this investigation. First, it was hypothesized that
gender would be a treatment moderator with studies
containing mostly males resulting in significantly lower
effect sizes than those containing mostly females. The
analyses conducted to explore this hypothesis did not result
in significant findings. For this analysis, a study had to
contain a disproportionately high number of the same sex in
its sample (i.e., over 2/3 male subjects or over 2/3 female
subjects). The studies containing mostly males had a mean
effect size of 0.48, and the effect size for the studies
containing mostly females was 0.66. The 95% confidence
intervals indicated these means were not significantly
different. Interestingly, the sixty-seven studies which
contained more equal numbers of males and females resulted
in an effect size of 0.53 which falls in between the other
two groups. This suggests that gender might be an important
factor to continue to examine in future research.

The second unsupported hypothesis was that different
sources of outcome measures would yield significant
differences. It was hypothesized that differences would
exist for measures completed by parents and children and
teachers and children; however, the analyses failed to find any significant differences. Once again, the mean effect sizes differed but not significantly when confidence intervals were examined. The average effect size for child self-report was 0.35; it was 0.25 for teachers, and 0.45 for parents. These results suggest that improvements appear to be perceived equally from these three sources of outcome data.

Despite the null findings from homogeneity analyses, a different result was obtained from a correlational analysis which selected only those studies for which dual reports were available for the sources of interest. The non-significant correlation between teacher and child reports (0.05) substantiated the hypothesis that agreement between these two sources would be low; however, the correspondence between parent and child outcome data was much higher (0.49). It makes sense that parents would be better able to sense changes in their children than teachers.

In both cases, the mean effect sizes for children were higher than those for parents or teachers; however, the discrepancy was larger for reports from teachers. Children seem to report more improvements as a result of treatment than either parents or teachers do. The fact that parents are more in tune with these changes than teachers might reflect the additional time parents spend with children, and the likelihood that parents have more opportunity to observe
therapeutic changes. For example, parents are more likely to be aware of changes in physiological functioning or changes in personality measures than teachers. Changes on behavioral measures which were used frequently in these studies might be more recognizable to teachers; however, behavioral measures may not be the most appropriate measures for assessing internalizing symptomatology. Therefore, teachers are probably not the best sources of outcome data for measuring certain kinds of functioning.

**Discussion of Descriptive Findings**

Several aspects of treatment studies for internalizing problems are noteworthy. First, it appears that the interventions contained in this review sampled a wide range of presenting problems, ages and types of therapists. For example, these interventions utilized different kinds of therapists such as professionals and paraprofessionals; they evaluated children ranging in age from three to eighteen and studied numerous types of presenting problems.

Second, this sample contained a range of treatments, and this range was consistent across different presenting problems. When comparing the treatment types found in this review to those found in other general reviews of child psychotherapy, it seems that research on the treatment of internalizing disorders samples more diverse types of treatments than other reviews have indicated (Weisz et al., 1987; Weisz et al., 1995). This review included more non-
behavioral studies (38.8%) than previous reviews which only contained 12-20% non-behavioral treatments (Weisz et al., 1987; Weisz et al., 1995). Non-behavioral types of treatments may be more appealing to clinical researchers since such treatments could address underlying feelings of distress which may be causing the internalizing disorder and would likely be ignored by behavioral treatments.

Third, there are several gaps in the literature which should be addressed by future studies of internalizing disorders. First, although 56 studies did include children with clinically relevant problems, another 56 did not provide enough information about the children being treated to assess this variable. Such information is extremely important for the generalizability of findings. Psychotherapy research on children has been criticized for its over-reliance on evaluating children who do not present with the same level of pathology as those typically referred for treatment (Kazdin, 1993; Weisz & Weiss, 1989). When studies do not describe the level of problem severity contained in their samples, it is difficult to determine how generalizable their findings are to actual clinic populations.

A second gap in treatment research pertains to the racial composition of the samples. Several authors have recently asserted how crucial it is for future treatment studies to evaluate interventions which include racial
minority samples and which test treatments designed to be culturally sensitive (Costantino, Malgady, & Rogler, 1994; Kazdin, 1993). This sample of studies reflects the tendency of researchers to either seldom include racial minorities in their studies or to not report the racial background of their subjects. In this sample, only 10.7% of the studies specifically reported having any minorities in their study and over 80% did not report the racial composition of their sample at all. Therefore, it is unknown if current treatments for internalizing problems are appropriate for and successful with culturally and ethnically diverse groups.

The overreliance on evaluating short-term interventions represents a third gap present in current studies. Almost 70% of the studies included in this review contained fewer than 20 sessions. Given that the sessions for most interventions were conducted on a weekly basis, the majority of these interventions involved a treatment period of less than six months. This is problematic for several reasons. First, clinical surveys have revealed that most children referred for treatment are seen for longer periods of time than the children studied in this review (Kazdin, Siegel & Bass, 1990; Tuma & Pratt, 1982). Therefore, the interventions evaluated in this review may not be generalizable to treatment as it occurs in clinic settings. Second, brief treatments may not be as successful for
children with multiple presenting problems. This interpretation is supported by the overall lower effect sizes generated by interventions treating children with multiple problems. Given these considerations, future studies need to consider evaluating longer-term treatments which might reflect the therapy typically conducted in clinic settings and which may be more effective for comorbid conditions.

Methodological Issues

Many of the studies in this review had sound methodological features. For example, 79% had random assignment to conditions; 88.7% had attrition rates of less than 10%; 65.2% used more than one outcome measure, and 61.2% included a measure to assess the generalized impact of treatment. Other design features which were more lacking included collecting follow-up data, using at least one normed outcome measure, and use of an attention placebo control group. These characteristics were only present in 32%, 30.2%, 27.9% of studies respectively. Results from the multiple regression indicated that these seven methodological features when entered together were effective predictors of outcome. In particular, inspection of the beta weights revealed that the collection of follow-up data, use of multiple outcome measures, and inclusion of a measure designed to assess the generalized impact of treatment are
the most important methodological criteria for predicting outcome.

Studies which collect follow-up data typically had larger effect sizes (0.55 vs. 0.38). This finding is difficult to explain. It is possible that researchers of studies which generated very positive results at post-treatment were more interested in examining the long-term impact of their interventions than those researchers with modest to poor outcome data.

The inclusion of multiple measures or inclusion of a measure assessing generalized impact was also correlated with lower effect sizes. These two findings are related due to the tendency for studies using multiple outcome measures to include a measure assessing the generalized impact of treatment. Results from this review suggest that it is more difficult to obtain generalization of treatment effects; therefore, it makes sense that studies with outcome measures which assess generalization would yield lower effect sizes.

Although these results suggest that including multiple and diverse measures will decrease the overall effect size, this should not discourage researchers from including them. Multiple measures provide information about how an intervention impacted different areas of functioning. This type of data can be very important for informing future research. In particular, measures of generalized impact are necessary for determining if the changes observed as a
result of treatment are clinically meaningful (i.e., if the changes have generalized to other areas of functioning other than the target area). Therefore, it is crucial that future research include multiple measures so that differences in outcome can be better understood in the context of child psychotherapy.

Limitations of the Current Review

While this review produced some important findings, it has some limitations. First, achieving homogeneity in all cells of every variable was not reached. Only two variables contained completely homogeneous cells. Most variables, including those which predicted outcome, did not obtain homogeneity in all cells. Therefore, some caution must be exercised in interpreting the findings from the analyses.

Small cell sizes for some variables represents another limitation. For example, the small number of studies treating depression (n=8) made interpretation of these findings difficult. In the case of the interaction analyses, the small cell sizes for problem type, treatment type and therapist type prevented many of the homogeneity analyses from being conducted. Therefore, any differences which might exist among the interactions of these variables could not be detected. Future studies would want to have larger n's so that more fine-grained analyses could be explored.

A final limitation of this review is that most of the variance attributable to effect sizes was not accounted for
in the multiple regression analyses. The variance accounted for was only 31%; therefore, the remaining 69% of unaccounted variance suggests a need to identify additional significant predictors of outcome.

**Future Directions**

The results from this investigation offer some guidance for future research. First, future research on the clinical treatment of internalizing disorders in children and adolescents should examine the effectiveness of treatment approaches for different presenting problems. Although the results of this review suggest that a variety of treatments can be effective for the treatment of internalizing disorders, it is still unclear which approaches are effective for which kinds of problems. Gaining a better understanding of treatment specificity (i.e., which treatment is best for which problem) will provide useful directions for future research as well as for informing clinical practice. We also need to increase the power of interventions for children with multiple problems. Perhaps successful treatment techniques for single presenting problems could be combined and administered to children who have multiple problems, and these interventions could be lengthened as well. The importance of future research evaluating treatments for children with multiple problems should be emphasized as a priority for treatment researchers given the prevalence of comorbidity in clinic populations.
The influence of methodological features represents another important implication for future research. The methodological variables measured in this study accounted for significant variance in determining outcome. This result points to the relevance of evaluating methodological variables in treatment research. Future studies should incorporate more rigorous standards into their designs. Treatment studies should meet certain methodological standards such as those measured in this review. In addition, features such as operational definitions, standardization procedures, and manipulation checks will be important as well for assessing internal validity. These features were included so infrequently in this sample of studies that no analyses of these features could be done. However, they represent important challenges for researchers who are designing future treatment studies.

Given that the model produced by this investigation was only able to account for 31% of the variance attributable to outcome, future treatment studies should focus on identifying additional variables which can enhance predictability of outcome. There are some variables which might have potential for predicting treatment outcome but were not measured in this review. For example, the differences found for age levels may suggest a need to give more consideration to children's developmental level when designing interventions.
A second focus for future research is the influence of family variables such as family functioning, marital distress, parental support or rejection of the child, and parental psychopathology. Each of these variables have the potential to influence treatment outcome. In addition, many reviewed studies were conducted in the schools and probably included only minimal contact with parents. Only 20 studies collected any outcome data from parents. Perhaps interventions might have been more successful had there been more contact with family members both during and following treatment.

A third possible variable which could not be measured relates to the level of pathology. Very few studies provided any information regarding formal diagnoses of treated children. While previous authors have suggested that many children in psychotherapy outcome research are not clinically distressed (Kazdin, 1993; Weisz & Weiss, 1989), the inclusion of information relevant to diagnostic criteria may be quite useful for issues of generalizability and for interpreting improvements in functioning. One possibility is for future studies to include two control groups: one group containing untreated children with similar problems and a second group containing children who do not have problems. This would enable researchers to draw more definite conclusions about the efficacy of their treatments.

In summary, child psychotherapy research has
demonstrated that internalizing disorders can be treated effectively. The positive findings from this review confirm other meta-analyses (Casey & Berman, 1985; Weisz et al., 1987; Weisz et al., 1995). Nevertheless, while the field of clinical child psychology has made considerable gains in the study of child treatment, there is still much more to be learned. In particular, research on the treatment of internalizing disorders has lagged behind the study of externalizing pathology. Findings from this review offer reason for optimism as well as some important directions for future research.
APPENDIX A

LIST OF JOURNALS INCLUDED IN THE MANUAL SEARCH PROCEDURES
APPENDIX A

List of Journals Included in the Manual Search Procedures

1. American Journal of Community Psychology
2. Behavior Modification
3. Behavior Research and Therapy
4. Behavior Therapy
5. Cognitive Research and Therapy
6. Elementary School Guidance and Counseling
7. Journal of Abnormal Child Psychology
8. Journal of Applied Behavior Analysis
10. Journal of Community Psychology
11. Journal of Consulting and Clinical Psychology
12. Journal of Counseling Psychology
14. Psychology in the Schools
15. School Counselor
APPENDIX B
CODING SCHEMA FOR META-ANALYSIS OF INTERNALIZING DISORDERS
APPENDIX B

Coding Schema for Meta-analysis of Internalizing Disorders

I. Study Characteristics:

1. Study ID# (001-999) 1-3
2. Year of Publication (code last 2 digits) 4-5
3. General Psychotherapy (check if applicable) 6
4. School-based (check if applicable) 7
5. Group (check if applicable) 8
6. Prevention (check if applicable) 9
7. Affective Education (check if applicable) 10
8. Source (1-5)
   1=published article
   2=book
   3=dissertation
   4=conference paper
   5=other
9. Total number of treatment groups 12-13
10. Total number of comparisons 14-15
11. Total number of outcome measures 16-17
12. Follow-up data available (1-2)
   1=yes
   2=no

II. Design Characteristics

13. Type of design (1-5)
   1=Pretest-Posttest with nonequivalent control group (NECG)
   2=Posttest only with NECG
   3=Randomized True Experiment
   4=Other (e.g. matching)
   5=not available
14. Group assignment procedure (1-6)
   1=random
   2=matching
   3=available intact
   4=voluntary self-selection
   5=other
6=not available

15. Total sample size-assigned  
(all treatment groups and control groups)  21-24

16. Total sample size-completed posttest  
(all treatment groups and control groups)  25-28

Not ascertainable? Code 00 (assume same as 9 if not stated)

17. Overall quality code

III. Subject Information

18. Number of males in total sample  29-31
Number unknown? Code 999

19. Mean age of subjects to nearest tenth year 32-34
Number unknown? Code 000

20. Ethnic sample characteristics (1-4)  35
1=majority or all white  
2=majority or all minority  
3=mixed  
4=unknown

21. Special sample characteristics (1-6)  36
1=retarded  
2=learning disabled  
3=underachievers  
4=hospital/dental patients  
5=other  
6=unknown  
7=none

22. Source of subjects (1-7)  37
1=clinical inpatients  
2=clinical outpatients seeking treatment  
3=volunteers for special project  
4=subjects chosen through problem-oriented observation, measurement, or recommendation  
5=hospital/dental patients  
6=convenient  
7=mixed/other  
8=unknown

22.5 General seriousness of problem (1-4)  38
1=none  
2=mild  
3=moderate to severe  
4=of uncertain nature/degree
5=at risk

23. Target Problem (1-16) 39-40
1=social isolate
2=fears/phobias
3=anxiety
4=enuresis
5=somatic problems
6=depression
7=other or mix of 1-6
(1-7 indicate internalizing symptomology)
8=impulsive/hyperactive
9=non-compliant/management problem/behavior problem
10=psychotic/autistic
11=other or mix of 8-10
12=social skills, undefined
(8-12 indicate externalizing symptomology)
13=mix of 1-12
14=none
15=unknown
16=other (academics, achievement)

24. Academic learning problems (1-3) 41
1=present
2=absent
3=unknown

IV. Therapist Characteristics

25. Number of therapists (code 00 if unknown) 42-43
26. Experience level of therapist (1-8) 44

1=mental health professionals (PhD in Psychology, social work; MD in Psychiatry; school guidance counselor)
2=professional trainees (graduate students in psychology, interns, practicum students, psychiatric residents)
3=parents
4=teacher
5=other non-professionals
6="experimenter"
7=mixed
8=unknown

V. Comparison Information

27. Comparison number 45-46
28. Type of Comparison (1-4) 47
1=treatment vs. control
2=behavioral vs. nonbehavioral
3=individual vs. group
4=combination

29. Type of Control Group (1-7) 48
1=none
2=no treatment (assume if not stated)
3=wait-list
4=attention-placebo
5=other
6=not available
7=mixed (i.e. 2&4)

30. Sample size of treatment group for this comparison 49-51
31. Sample size of control group for this comparison 52-54

VI. Treatment Characteristics

32. Type of treatment (1-4) 55
1=behavioral
2=nonbehavioral
3=mixed
4=unknown

33. Method of delivery (1-4) 56
1=individual
2=group
3=mixed
4=unknown

34. Number of treatment sessions (code 00 for unknown) 57-59

35. Average length of treatment sessions in minutes (code 999 for unknown) 60-62

36. Treatment setting (1-9) 63
1=school
2=home
3=mental health, community mental health or psychology/psychiatry clinic
4=general hospital or dental clinic
5=residential treatment center (psychiatric or special school)
6=camp
7=combination of at least two of the above
8=other
9=unknown
VII. Characteristics of Outcome Measures

37. Type of outcome measure (1-9)
   1=independent behavioral observation
   2=nonindependent behavioral observation
   3=peer sociometric
   4=normed rating scale or behavioral checklist
      (or psychometrically adequate - someone else has used it before)
   5=nonnormative/experimenter constructed instrument
   6=achievement test or intellectual measure
   7=other performance measure (e.g. MFF)
   8=school grades
   9=objective performance measure (e.g. days in school, arrests, approaching feared object)

37.5 Specific or generalized impact of treatment
   1=specific
   2=generalized

38. Source of outcome measure (1-10)
   1=independent observers
   2=parents
   3=therapist
   4=teachers/school
   5=peers
   6=subject self-report
   7=subject performance measure (on an achievement, IQ, or cognitive measure)
   8=other (expert judges, not independent observers, or therapists, or 1-7)
   9=mixed
   10=unknown

39. Dimension of adjustment (1-11)
   1=fear/anxiety
   2=cognitive skills
   3=global adjustment
   4=social adjustment/social skills
   5=achievement
   6=personality
   7=交接-自尊
   8=bed-wetting
   9=mixed
   10=unknown
   11=physiology

39.5 Type of adjustment or change measured (1-9)
   1=behavioral
   2=personality
   3=academic performance
   4=sociometric
5 = cognitive tempo
6 = cognitive problem-solving skills
7 = physiological measure
8 = other
9 = mixed

### VIII Effect Size Information

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>40.</td>
<td>Reliability of measure</td>
<td>71-74</td>
</tr>
<tr>
<td></td>
<td>(code 9999 if not available)</td>
<td></td>
</tr>
<tr>
<td>41.</td>
<td>Effect size at posttreatment</td>
<td>76-79</td>
</tr>
<tr>
<td></td>
<td>(if NA then +9999)</td>
<td></td>
</tr>
<tr>
<td>42.</td>
<td>Length of follow-up in weeks</td>
<td>80-82</td>
</tr>
<tr>
<td></td>
<td>(if NA code then 000)</td>
<td></td>
</tr>
<tr>
<td>43.</td>
<td>Effect size at follow-up</td>
<td>84-87</td>
</tr>
<tr>
<td></td>
<td>(if NA then code +9999)</td>
<td></td>
</tr>
<tr>
<td>44.</td>
<td>How effect size was calculated (1-14)</td>
<td>88-89</td>
</tr>
<tr>
<td></td>
<td>1 = means/standard deviation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 = anova summary table</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 = t score</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 = raw data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 = ANCOVA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6 = probit or chi square/nonparametric</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7 = change scores</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 = estimate from p</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9 = correlations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 = nonsignificant and no statistical info</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11 = Holmes method</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12 = posttest adjustment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13 = mixed (two or more of above were used)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>combination of methods</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14 = separate ES methods for post &amp; follow-up</td>
<td></td>
</tr>
<tr>
<td>45.</td>
<td>Source of data (1-3)</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>1 = standard information provided</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 = data drawn from graphs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 = 2 week test-retest reliability used</td>
<td></td>
</tr>
<tr>
<td></td>
<td>with change scores</td>
<td></td>
</tr>
<tr>
<td>46.</td>
<td>Number of this outcome measure</td>
<td>91-92</td>
</tr>
<tr>
<td>47.</td>
<td>Measure to be combined with others</td>
<td>93-94</td>
</tr>
<tr>
<td></td>
<td>(start numbering consecutively, collecting all</td>
<td></td>
</tr>
<tr>
<td></td>
<td>similar measures with the same number)</td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


VITA

The author, Kathryn Ann McGlinchey is the daughter of Terry F. and Carol A. (Fehn) McGlinchey. She was born on July 8, 1971 in Indianapolis, Indiana.

Ms. McGlinchey's elementary education was obtained in Indianapolis, Indiana. Her secondary education was completed in 1989 at Cathedral High School in Indianapolis, Indiana.

In May, 1993, Ms. McGlinchey completed her undergraduate education at Loyola University Chicago. She received a Bachelor of Science Honors degree in Psychology, Summa Cum Laude. While attending Loyola University Chicago, Ms. McGlinchey was a member of Alpha Sigma Nu Honor Society and Psi Chi Honor Society.

In August, 1994, Ms. McGlinchey began her graduate studies towards her Ph.D. in Clinical Psychology at Loyola University Chicago. She completed her Master of Arts degree in May, 1996.
APPROVAL SHEET

The thesis submitted by Kathryn Ann McGlinchey has been read and approved by the following committee:

Dr. Joseph A. Durlak
Professor, Psychology
Loyola University

Dr. Paul Jose
Associate Professor, Psychology
Loyola University

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 incorporated 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.

Date  
Director's Signature

April 2, 1981  
Joseph A. Durlak