Discovering How Youth Psychotherapies Work: Three Approaches to Identifying Mediators of Treatment Outcome

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citable link</th>
<th><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:39987872">http://nrs.harvard.edu/urn-3:HUL.InstRepos:39987872</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Discovering How Youth Psychotherapies Work:
Three Approaches to Identifying Mediators of Treatment Outcome

A dissertation presented
by
Mei Yi Ng
to
The Department of Psychology

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Psychology

Harvard University
Cambridge, Massachusetts
June 2017
Discovering How Youth Psychotherapies Work:
Three Approaches to Identifying Mediators of Treatment Outcome

Abstract

Although a lot is known about which psychotherapies work, surprisingly little is known about how they work. Understanding the processes and mechanisms through which evidence-based psychotherapies (EBPs) improve outcomes can enrich theories of change and inform treatment refinement. I conducted three studies intended to identify candidate change processes and mechanisms that mediate the effects of EBPs for youths, and also to illustrate the use of quantitative methods that can advance research on change processes and mechanisms, but are rarely used in psychotherapy research.

In Study 1, I conducted a systematic review and meta-analysis of mediation effects in EBPs for youth depression. Using meta-analytic structural equation modeling, negative thinking, and social skills and relations were identified as robust mediators, and family dysfunction was not a significant mediator, across multiple randomized trials. Problem solving, pleasant activities, cognitive skills, and avoidant coping also did not significantly mediate outcomes, but analyses involved fewer trials; more data are needed on these candidate mediators as well as others (e.g., role engagement, therapeutic alliance) that had insufficient data for analysis.

In Study 2, I examined youth coping and therapeutic alliance as candidate mediators in a randomized effectiveness trial comparing cognitive behavioral therapy and usual care for youths with internalizing disorders. Bootstrapping of indirect effects with multiply imputed datasets revealed that improved youth-reported secondary control coping (i.e., adjusting oneself to adapt
to existing conditions) may be a potential change mechanism underlying the effectiveness of usual care, and improved parent-reported youth secondary control and primary control coping (i.e., influencing existing conditions to fit one’s goals) may be common factors predicting better outcomes across therapies.

In Study 3, I investigated whether parent-identified “top problems”—a measure of youths’ achievement of personalized treatment goals—as a candidate change process that might mediate overall reduction of youth internalizing and externalizing symptoms in a randomized effectiveness trial comparing modular EBPs to standard EBPs. Bivariate latent change score models suggested that top problems did not predict subsequent symptom change, but that symptoms predicted subsequent change in top problem severity. Thus top problems was not supported as a change process through which modular EBPs outperformed standard EBPs.
Study 3: Clarifying the Direction of Change Between Personalized “Top Problems” and Symptom Outcome Trajectories in Modular vs. Standard Evidence-Based Psychotherapies for Internalizing and Externalizing Youths

Abstract

Introduction

Method

Results

Discussion

References
Acknowledgements

I am truly grateful to have been supported by so many people in the completion of this dissertation and in my graduate school career. Dr. John R. Weisz is the best advisor I could have asked for; I have benefited immeasurably from his brilliance, wisdom, and nurturance. Drs. Katherine E. Masyn, Mark W. Lipsey, and Matthew K. Nock have been incredibly generous with their expertise and time—it is a real privilege to have them on my committee. Dr. David A. Langer was a wonderful clinical supervisor and has been a trusted colleague since my first year as a graduate student; I am delighted that he will be the outside examiner for my dissertation defense. In addition, I am fortunate to have received valuable assistance from many past and present volunteers and members of the Harvard Youth Mental Health Laboratory—most especially from Katherine (Kate) A. DiVasto, Samantha Cootner, and Nazc-a-ru (Aru) H. Gonzalez—who were instrumental to the completion of Study 1. Dr. Mike W.-L. Cheung and the Harvard Institute for Quantitative Social Science statistical consultants provided much needed guidance on data analysis and management; and many study authors kindly retrieved and shared their data with me, which was often no small burden! Moreover, I am extremely lucky to have Drs. Marina Tolou-Shams, Dale McNiel, and others, as my internship supervisors, who gave me the time I needed to complete this dissertation during a professionally and personally challenging year, as well as a cohort of interns and postdocs at UCSF who were ever willing to help. I really hit the jackpot with Dr. Nancy Lau and Dr. Christine B. Cha, my best friends and study buddies who have continued to provide comfort and counsel after they graduated. Finally, I am indebted to my parents, Ng Chee Leong and Sandra Wong Tuck Chuen, my siblings Ng Mei Fay and Ng Kin Foong, and Dr. William K. Lamb, for their unconditional, undying love.
Introduction

We know a lot about which psychotherapies work. Expert workgroups in psychological science have reviewed the research literature to identify psychotherapies that show evidence of beneficial effects. Their findings are documented in numerous media, including academic journals (Chambless et al., 1996, 1998; Kendall, 1998; Lonigan & Elbert, 1998; Silverman & Hinshaw, 2008), books (Nathan & Gorman, 1998, 2002, 2007), and websites (Association for Behavioral and Cognitive Therapies & Society of Child Clinical and Adolescent Psychology, n.d.; Center for the Study and Prevention of Violence, n.d.; Society of Clinical Psychology, n.d.; Substance Abuse and Mental Health Services Administration, 2016). Although evaluation criteria for psychotherapies have varied across workgroups and evolved over time (see Southam-Gerow & Prinstein, 2014), all criteria were written with the common goal of identifying psychological interventions that demonstrated beneficial effects when compared to a control group or alternative treatment. The interventions that meet these criteria are referred to as evidence-based psychotherapies (EBPs; Weisz & Kazdin, 2017).  

With such extensive efforts to identify what works, it is surprising how little is known about how psychotherapies work (Kazdin, 2007). When I proposed this dissertation, I was aware of no journal special issues, books, or websites devoted to comprehensive searches of the literature to find the processes and mechanisms through which psychotherapies produce desired change in outcomes that are supported by evidence. Since then, one book (Maric, Prins, & Ollendick, 2015) reviewing mediators (and moderators) in youth psychotherapy has been

---

1 No date was provided on the source; “n.d.” is listed in place of the year according to publication manual of the American Psychological Association (2009).

2 Interventions that meet efficacy criteria have also been termed empirically validated therapies (Chambless et al., 1998), empirically supported therapies (Chambless & Hollon, 1998), and evidence-based (psychosocial) treatments (Silverman & Hinshaw, 2008), and have subtly different connotations. I use evidence-based psychotherapies as a general term to represent all these terms.
published. This development suggests, first, that researchers are paying more attention to psychotherapy change processes and mechanisms; and second, that there is an accumulating pool of evidence on how psychotherapy works that needs to be systematically located, organized, and summarized. This increasing attention to change processes and mechanisms is encouraging, given that a deeper understanding of how psychotherapies help individuals achieve desired change could inform theories of behavior change and guide the practice of psychotherapy (Kazdin & Nock, 2003).

**Background and Terminology Related to the Study of How Psychotherapies Work**

To sharpen the conceptual clarity of my ideas and to improve the precision of my writing, I discuss in this section key conceptual and statistical terms used in the literature on how psychotherapy works, noting areas of overlap and distinction. Doing so necessarily entails providing some historical background into researchers’ use of terminology, and how the terms and their definitions have evolved over time. I close this section with an explanation of the terms I use and what they mean in this dissertation.

Traditionally, efforts to study how psychotherapies work have come under the umbrella of *process-outcome* research—a large body of literature dating back to the 1950s, and estimated to encompass more than 2,000 published studies (Orlinsky, Rønnestad, & Willutzki, 2004). Orlinsky et al. (2004, p. 311) define *psychotherapy or treatment process* as “specific happenings and events observed in therapy, including (primarily) the actions, experiences, and relatedness of patient and therapist in therapy sessions when they are physically together, and (secondarily) the actions and experiences of the participants specifically referring to one another that occur outside of therapy sessions when they are not physically together.” Examples offered include the therapeutic alliance, patient expressiveness, and therapy homework completion. Orlinsky and
colleagues (p. 313) also differentiated treatment process from change process, the “change occurring in aspects of the client’s psychological functioning (e.g., manner of experiencing, construal of meaning…)” that presumably cause outcome to improve. Process–outcome research has predominantly focused on common factors shared by psychotherapies of various theoretical orientations, and has largely employed correlational studies of naturalistic psychotherapy (Crits-Christoph, Connolly Gibbons, & Mukherjee, 2013). When the process–outcome correlation is significant, the process may be called a predictor of treatment outcome (e.g., Kraemer, Wilson, Fairburn, & Agras, 2002), but the correlational design precludes tests of whether the psychotherapy caused the process, and whether the process caused the outcome.

The rise of the EBP movement in the 1980s established the randomized controlled trial (RCT) as the gold standard design for psychotherapy research (Chambless & Hollon, 1988; Silverman & Hinshaw, 2008, Southam-Gerow & Prinstein, 2014). In line with EBP criteria, psychotherapies were developed from specific theoretical models for particular disorders, standardized in manuals, and tested in RCTs (Crits-Christoph et al., 2013). The experimental manipulation of treatment condition allowed inference of whether treatment had causal effects on outcome. With the identification of EBPs came empirical efforts to validate the theories that purported to explain how they worked, and a shift towards the term, mechanism. Weersing and Weisz (2002) authored a comprehensive review of mechanisms of action for EBPs of youth, such as cognitive distortions in cognitive behavioral therapy (CBT) for depression, arousal in CBT for anxiety, and parent commands in behavioral parent training for conduct problems. They suggested that mechanisms refer to “processes involved in psychopathology” that are manipulated by psychotherapy “in the service of alleviating symptoms in already impaired

3 Although single-case experimental designs also permit causal conclusions to be drawn, and were admissible for meeting original EBP criteria (Chambless & Hollon, 1988), psychotherapy researchers have gravitated towards RCTs (Barlow & Nock, 2009).
youth” (Weersing & Weisz, 2002, p. 4). This conceptualization of mechanisms is similar to Orlinsky et al’s (2004) definition of change processes.

The ascendance of the EBP movement coincided with the emergence of mediation analysis in psychological research, following publication of Baron and Kenny’s (1986) seminal paper contrasting statistical moderation and mediation. Baron and Kenny (see also Judd & Kenny, 1981) defined a mediator as a variable that accounts for the relationship between the predictor and outcome, and thus represents the mechanism through which the predictor impacts the outcome. Researchers (e.g., Holmbeck, 1997; Kraemer et al., 2002; Weersing & Weisz, 2002) began advocating for hypothesized psychotherapy processes or mechanisms to be tested as mediators of the treatment condition–outcome relationship in RCTs. The experimental design of the RCT offers the advantage of testing whether the psychotherapy caused changes in the outcome and in the hypothesized process or mechanism, although the tested relationship between the hypothesized process or mechanism and outcome remains a correlational one.

Mediators had so often been construed as psychotherapy processes or mechanisms that Kazdin and Nock (2003; see also Kazdin, 2007) tried to disentangle the two by emphasizing the statistical function of mediators, and defining mechanisms of change as “processes or events that lead to and cause therapeutic change” (p. 1117) and that must meet multiple methodological and logical criteria, beyond being a significant mediator. These criteria, adapted from the scientific and medical literature on establishing causation (e.g., Hill, 1965) for psychotherapy research, include strength, consistency, and specificity of the mechanism–outcome relationship; a dose-response relationship between mechanism and outcome; plausibility and coherence of the mechanism; experimental manipulation of the candidate mechanism itself; and temporal precedence of the candidate mechanism in relation to the outcome through frequent assessments.
of both mediator and outcome. Kazdin and Nock’s examples of therapeutic alliance and negative cognitions suggest that their definition of mechanisms of change might encompass psychotherapy processes and change processes as conceptualized by Orlinsky and colleagues (2004), and also mechanisms of action as conceptualized by Weersing and Weisz (2002).

Finally, Doss (2004) created a conceptual model that distinguishes among these interrelated constructs. He categorized what the therapist does during sessions as therapy change processes (e.g., specific intervention strategies used by therapists), what the client does during sessions or as a result of the sessions as client change processes (e.g., client completion of homework assigned by the therapist), and changes in client characteristics and skills outside of sessions as change mechanisms (e.g., client coping strategies). In Doss’ model, therapy and client change processes are thought to influence one another bidirectionally; change processes are then hypothesized to have causal effects on change mechanisms, which in turn cause changes in outcomes (though events outside of therapy could also affect client change processes, therapy processes may not actually influence the hypothesized client change process, and change mechanisms could influence therapy change processes). Consistent with Kazdin and Nock (2003), Doss emphasized the importance of clarifying timelines for change processes, change mechanisms, and outcomes, and of experimental design in determining causality. However, Doss diverged from Kazdin and Nock in his assertion that change processes, but not change mechanisms, can be experimentally manipulated as only change processes can be directly controlled by the therapist. To illustrate this idea, the therapist has complete control over the therapy change process of whether to use an intervention technique from the cognitive therapy model for depression; and she can guide the client to identify negative thoughts and come up with alternative responses in-session or as part of homework, thus she has partial control over
this client change process. However, she has no direct control over the client’s frequency and intensity of negative cognitions in everyday life, or how effectively he identifies and modifies his negative cognitions on a regular basis. The therapist might indirectly influence the level of these change mechanisms through what she does in session and assigns for homework, but there is likely to be between- and within-client variation in whether, and to what extent, she can get this mechanism to change. Thus Doss recommended mediation analysis rather than experimental manipulation as the appropriate method of studying change mechanisms.

Given this dizzying array of conceptually overlapping, similarly worded terms involved in studying how psychotherapy works, it would be prudent to clarify the terms that I use in this dissertation. I employ therapy change process, client change process, and change mechanism as defined in Doss’ (2004) model, consistent with recent usage of terminology by other psychotherapy researchers (e.g., Crits-Christoph et al., 2013). Differentiating these specific aspects of therapeutic change that occur after assignment or start of a treatment, and lead to improvement in outcome, will allow me to convey my ideas more precisely. In addition, I use the term therapy–client change process for constructs such as the therapeutic alliance that combine elements of the therapy change process and client change process, and change process as an overarching term for any kind of change process. I also consider change mechanism and mechanism of change to be equivalent and use them interchangeably. A candidate, hypothesized, or putative process or mechanism is thought to be a change process or mechanism but has not met all the methodological and logical criteria delineated by Kazdin and Nock (2003) and Doss (2004). In contrast to the substantive meaning of the aforementioned set of terms, I use predictor and mediator as statistical terms to denote, respectively, a variable that is associated with an outcome and a variable that accounts for the treatment condition–outcome association in the
context of an RCT. I refer to variables that will be tested as predictors or mediators as a predictor or mediator variable, or as a candidate, hypothesized, or putative predictor or mediator.

Significant statistical prediction and mediation will be considered one of the methodological criteria for change process or change mechanism status.

**Overview of Three Dissertation Studies**

The primary aim of my dissertation is to identify mediators of outcome in EBPs for children and adolescents (herein, “youths”) as an initial step towards understanding change processes and mechanisms in youth psychotherapies. Establishing even one change process or mechanism will undoubtedly be a complex, resource-intensive endeavor that will encompass multiple studies with a range of designs, samples, conditions; plausible comparison candidate processes and mechanisms; and frequent assessments. Focusing on the therapies with a large evidence base, and drawing on the extant RCT data that make up this evidence base, is one way to inform future research efforts. Significant mediators identified in my dissertation through secondary analysis of RCT data could be promising candidates that are more likely to bear fruit (Ng & Weisz, 2016; Weisz, Ng, Rutt, Lau, & Masland, 2013).

A secondary aim of my dissertation is to explore and apply methodological approaches that offer distinct advantages in the search for candidate processes and mechanisms. Aside from the frequent assessments required to clarify temporal relationships, there is the issue of responsiveness—the therapist often adjusts the use of certain techniques or practices depending on the patient’s response, thus the amount and type of therapy change process used can both cause, and result from, changes or lack of change in client change processes, mechanisms, or outcomes (Stiles, Honos-Webb, & Surko, 1998; see also Crits-Christoph et al., 2013; Doss, 2004). Other obstacles include the low power of popular mediation tests (Fritz & MacKinnon,
2007), compounded by RCT sample sizes that are powered to detect between-group treatment effects but not mediation effects (Kraemer et al., 2002); as well as the restrictive assumptions of popular mediation tests. These considerable methodological challenges have likely contributed to the dearth of knowledge on change processes and mechanisms. I have thus attempted to demonstrate the use of three statistical approaches that mitigate some of these challenges.

Study 1 is a systematic review and meta-analysis focused on mediation effects in EBPs for youth depression, with the aims of identifying robust mediators in the expected direction, robust mediators in the unexpected direction, variables that do not reliably mediate outcomes, and variables that have not been adequately assessed as mediators. The systematic review portion updates and expands on prior reviews, helps make sense of existing findings and trends in the use of mediation approaches over time, and charts future directions for research. The meta-analysis portion was conducted using a recently developed meta-analytic structural equation modeling approach, and estimated the magnitude and direction of mediation effects for seven candidate mechanisms that have been assessed in RCTs testing EBPs for youth depression. This pioneering mediation meta-analysis is intended to serve as a working model for how the field might exploit largely untapped existing data to inform more focused research on change processes and mechanisms.

Study 2 examines candidate change processes and mechanisms through which manualized cognitive behavioral therapy (MCBT) and usual clinical care (UC) work under everyday clinical conditions to ameliorate internalizing problems (anxiety and depression) among youths. Using data from two randomized effectiveness trials that found comparable outcomes for MCBT and UC, I examined youth coping and therapeutic alliance as mediators of the effects of treatment condition, cognitive strategies, and behavioral strategies on posttreatment
internalizing symptoms. Multiple imputation was conducted to address missing data, and bootstrapped confidence intervals were constructed to assess mediation effects and combined across the multiply imputed datasets. Both these procedures are “gold-standard” methods that offer several advantages over alternatives but are rarely used together.

**Study 3** investigates a candidate change process using data from a randomized effectiveness trial with weekly assessments that allows clarification of temporal relationships between the candidate process and outcome. This trial found that a modular therapy comprising EBP practices outperformed standard EBPs and UC among youths with elevated internalizing or externalizing symptoms, or some combination thereof. The candidate mediator was “top problems” identified by parents, which served as a measure of youths’ achievement of personalized treatment goals. The outcome was internalizing and externalizing symptoms, assessed with a standardized, nomothetic measure. Using a bivariate latent change score approach, I assessed whether changes in top problem severity predicted, and were predicted by, changes in symptom severity; and whether the predictor variable(s) also mediated the effects of treatment condition on the other variable over 16 weeks of treatment. This is one of the few intensive longitudinal studies, to date, of candidate change processes in youth psychotherapy research.
References


Dissertation Study 1:
How do Evidence-Based Psychotherapies for Youth Depression Work?

Synthesizing Mediation Effects from 30 Years of Evidence

Mei Yi Ng
Harvard University
Abstract

**Background:** Therapy for youth depression produces only small–medium effects—markedly smaller than effects with other problems (Weisz et al., 2006, 2017). Understanding how evidence-based psychotherapies (EBPs) for youth depression work may guide efforts to improve their efficacy. To accelerate the identification of change processes and mechanisms for youth depression, we conducted (a) a systematic review, and (b) a meta-analysis of mediation effects.

**Method:** Drawing from a large database of youth psychotherapy RCTs (Weisz et al., 2017), we identified 40 RCTs that assigned 5,608 depressed youths (mean age=14.0 years, 39.9% male) to 50 EBP and 44 control conditions. We conducted literature searches and contacted authors to locate 123 articles and 9 dissertations or theses published 1982–2014 that report data from the 40 RCTs. We coded participant demographics, candidate mediator (CM) type, and correlations among treatment condition, CMs, and outcomes at pre-, mid-, posttreatment, and follow-up.

**Results:** (a) Our systematic review revealed that 75% of RCTs measured CMs, predominantly at pre- and posttreatment. However, only 15% analyzed CMs as mediators, and these yielded mixed or nonsignificant findings. (b) Meta-analytic structural equation modeling (MASEM) identified negative thinking and social skills and relations as significant mediators; pleasant activity engagement was a predictor; and family dysfunction, cognitive skills, and avoidant coping were neither predictors nor mediators of outcome. **Conclusions:** CMs are often measured but rarely tested as treatment mediators. MASEM identified negative thinking and social skills and relations as robust mediators across RCTs—these may be the most promising candidates to investigate as change mechanisms.

**Keywords:** youth, depression, evidence-based psychotherapies, mediators, meta-analysis
Study 1: How do Evidence-Based Psychotherapies for Youth Depression Work?

Synthesizing Mediation Effects from 30 Years of Evidence

Over the past 50 years, researchers have made remarkable progress in learning what psychotherapies work for which disorders in children and adolescents (herein, “youths”). These evidence-based psychotherapies (EBPs)—psychotherapies that have demonstrated efficacy in scientifically credible studies—have been identified and catalogued in books (e.g., Weisz & Kazdin, 2017), journal special issues or sections (e.g., Southam-Gerow & Prinstein, 2014), and websites (e.g., ABCT & SCCAP, n.d.). Among these EBPs are cognitive behavioral therapy (CBT) and interpersonal therapy (IPT) for youth depression (David-Ferdon & Kaslow, 2008; Weersing, Jeffreys, Do, Schwartz, & Bolano, 2016). CBT and IPT are among frontline interventions recommended for moderate to severe youth depression (Birmaher & Brent, 2007; National Institute for Clinical Excellence, 2015)

Despite the emergence of EBPs for youth depression, there is still much room for improvement. The most comprehensive meta-analysis of youth psychotherapies to date (Weisz et al., 2017) revealed that psychotherapies for youth depression produce on average only small-medium between-group treatment effects ($d = 0.29$)—significantly smaller than treatment effects for youth anxiety ($d = 0.61$) and conduct problems ($d = 0.46$). Similarly modest effect sizes ($d = 0.16$ to $0.34$) were found in earlier meta-analyses focused on psychological interventions to prevent or treat youth depression (Horowitz & Garber, 2006; Merry et al., 2011; Stice, Shaw, Bohon, Marti, & Rohde, 2009; Weisz, McCarty, Valeri, 2006). Moreover, the Treatment for Adolescents with Depression Study (TADS), a large multisite randomized controlled trial (RCT), found no significant difference between the effects of CBT and pill placebo, with serotonin reuptake inhibitor (SSRI) and combination CBT plus SSRI superior to both CBT and pill placebo
(TADS Team, 2004). These weak effects, taken together with the reported risks of SSRIs (Hetrick, McKenzie, Cox, Simmons, & Merry, 2012), and the devastating effects of depression on youth development (Rice, Lifford, Thomas, & Thapar, 2007), justify increased efforts to improve the efficacy of psychotherapy for youth depression.

**Studying Change Processes and Mechanisms**

Increasingly, there have been calls for such efforts to be guided by deeper understanding of the change processes and mechanisms through which treatments produce desired change, so that activation of these change processes and mechanisms may be intensified (Doss, 2004; Kazdin & Nock, 2003; Kraemer, Wilson, Fairburn, & Agras, 2002). In fact, investigating how treatments work has been championed by Kazdin and Nock (2003, p. 1117) as “probably the best short-term and long-term investment for improving clinical practice and patient care.” According to Doss (2004), such investigations may entail examining any of three related but distinct constructs that ultimately lead to improved symptom or functional outcomes: (a) *therapy change processes* refer to what the therapist does during sessions, such as therapist use of behavioral activation strategies or cognitive reframing techniques with the client; (b) *client change processes* refer to what the client does during or as a result of sessions, such as client completion of specific therapy homework tasks to engage in a scheduled activity or their ability to identify negative automatic thoughts and modify them to more rational, adaptive thoughts in session; and (c) *change mechanisms* refer to changes in client characteristics and skills outside of sessions on a daily basis, such as client everyday engagement in, and enjoyment of pleasant activities or their overall level of negative thinking or frequency of cognitive reframing skill usage. We note that “processes” and “mechanisms” are used interchangeably by many authors, sometimes to refer to what Doss calls therapy or client change processes, sometimes to refer to what Doss calls change mechanisms.
mechanisms, and sometimes to both. To improve the clarity of our ideas, we distinguish among the three constructs following Doss’ model, using the following heuristic: if the actions, thoughts, feelings, skills, experiences, or preferences belong to the therapist, then we labeled them therapy change processes; if the above belong to the client and applies only in session or in the context of ongoing treatment (e.g., therapy homework completion can only be measured if the client is receiving treatment), then we labeled them client change processes; if the above applies to the client regardless of whether they are receiving treatment or not (e.g., everyday engagement in pleasant activities can be measured in any individual, including waitlisted participants), then we labeled them change mechanisms or mechanisms of change (which we use interchangeably). Constructs that do not fall clearly into one of the categories were not addressed by Doss. For example, the therapeutic alliance is a relationship belonging to both the therapist and the client. We have thus labeled this a therapy-client change process. For brevity, we refer to therapy change processes and client change processes collectively as change processes when we are not distinguishing between them.

Enthusiasm for identifying change processes and mechanisms has grown in parallel with the sophistication of study designs, statistical methods, and logical criteria that can be used to evaluate them. Earlier research often assessed candidate processes as predictors of outcome, and candidate mechanisms as a type of outcome alongside symptom severity and adaptive functioning on which the main effect of treatment condition is tested (Crits-Christoph, Connolly Gibbons, & Mukherjee, 2013; Doss, 2004; Weersing & Weisz, 2002). Subsequently, mediation analysis became the recommended method for assessing candidate processes and/or mechanisms (Doss, 2004; Holmbeck, 1997; Kraemer et al., 2002; Weersing & Weisz, 2002). Mediation tests, in the context of RCTs, demonstrate whether the treatment changed levels of the candidate
process or mechanism, whether the candidate process or mechanism levels were in turn associated with outcomes, and whether the relationship between the candidate process or mechanism and outcome accounts for the treatment–outcome relationship. Therefore mediation tests are more informative and rigorous than the earlier approaches. Furthermore, mediation methods have advanced considerably from Baron and Kenny’s (1986; see also Judd & Kenny, 1981) causal steps approach to newer, improved approaches with higher power, fewer assumptions required, and broader applicability to a range of study designs and situations, such as significance testing of the conjoint mediational pathway (i.e., the indirect effect) rather than testing one path at a time, and using bootstrapping rather than normal distributions to generate confidence intervals of the indirect effect (Hayes, 2009). The increasing confusion and conflation of mediators, a statistical construct, with the substantive concepts of change processes and mechanisms, prompted further recommendations that multiple criteria need to be satisfied across multiple studies in order to determine whether a construct could be a change process or mechanism (Kazdin, 2007; Kazdin & Nock, 2003). These criteria include comparison of plausible candidate processes and mechanisms, frequent assessments to establish temporal precedence of change processes and mechanisms over outcomes, and experimental manipulation to demonstrate causal relationships

**Systematic and Meta-analytic Reviews of Change Processes and Mechanisms**

Undoubtedly, longstanding, resource-intensive research programs will be needed to satisfy the abovementioned criteria for even a single change process or mechanism; such programs should therefore be guided strategically by thorough knowledge of extant research. Yet the current literature poses considerable challenges to researchers looking for this guidance. Research reviews commonly focus on just one type of change process or mechanism, such as
therapeutic relationship variables (e.g., Karver, Handelsmann, Fields, & Bickman, 2006; McLeod, 2011), thus limiting comparisons among different candidate processes or mechanisms. Others devote a relatively brief section to reviewing evidence on change processes and mechanisms, instead focusing predominantly on treatment outcomes or other topics (e.g., David-Ferdon & Kaslow, 2008; Weisz, Ng, & Lau, 2015). Most do not report comprehensive, systematic search and screen procedures (e.g., Webb, Auerbach, & DeRubeis, 2012; Weersing, Rozenman, & Gonzalez, 2009; Weersing, Schwartz, & Bolano, 2015; Weisz, Ng, Rutt, Lau, & Masland, 2013), thus potentially leaving out portions of the evidence base, particularly null findings that may be reported but not emphasized, resulting in conclusions that may be biased.

Because systematic and meta-analytic reviews are the gold standard research design for taking stock of a body of literature (Hunt, 1997), we focused on these types of reviews. We undertook our own systematic search for systematic reviews and meta-analyses of change processes or mechanisms in therapies for youth depression, screening 830 abstracts from PsycINFO and PubMed (DiVasto, 2016), in addition to looking up reference trails and works of youth depression treatment researchers. We identified only two systematic reviews and one meta-analysis.

Weersing and Weisz (2002) conducted an extensive systematic review of RCTs testing EBPs targeting various youth problems; among the 13 RCTs of CBT and 1 RCT of IPT for youth depression identified, 12 assessed candidate mechanisms (e.g., cognitive distortions, self-concept, social adjustment, pleasant activities), of which 11 found that the EBP improved levels of the candidate mechanism compared to controls, one found that the candidate mechanism predicted outcome, 1 found that the candidate mechanism did not predict the outcome, and one
tested the candidate mechanism and found that it was not a mediator. Weersing and Weisz did not include change processes in their review.

The depression piece of the systematic review was updated by Weersing, Jeffreys, Do, Schwartz, and Bolano (2016), but narrowed in scope to include only change mechanisms that were tested as mediators in RCTs where samples met diagnosis for a depressive disorder or clinical cutoffs on standardized measures for depressive symptoms, or were identified by a mental health provider as experiencing clinical levels of depression symptoms. They located five RCTs of CBT (and none of IPT) that reported formal mediation tests, but findings were mixed for cognitive (n=5) and behavioral (n=2) candidate mechanisms, and the motivational (n=1) candidate mechanism was a significant mediator not for CBT alone, but for CBT plus SSRI combination treatment, compared to placebo.

Finally, Chu and Harrison’s (2007) meta-analysis of 14 RCTs demonstrated that CBT had significant larger-than-medium effects on symptom outcomes (d=0.60), small-to-medium effects on cognitive candidate mechanisms (d=0.35), and nonsignificant effects on behavioral (d=0.01) and coping (d=0.05) candidate mechanisms. They also noted that only three of these RCTs examined treatment mediators; however, they did not examine the relationship between candidate mechanisms and outcomes—a critical path in the mediation model. Chu and Harrison (2007) also did not review change processes.

These reviews provide valuable information, but also leave important questions in their wake. To summarize, there was a growing body of RCT literature published up till the mid-2000s indicating that CBT changed cognitive candidate mechanisms in the expected direction for depressed youth, relative to a control condition, but effects on behavioral and coping candidate mechanisms are equivocal, and associations between changes in candidate mechanisms and
changes in outcomes were rarely assessed. Moreover, only a minority of RCTs published up until 2014, all testing CBT, involved examining candidate mechanisms as mediators; they produced inconsistent results and failed to establish temporal precedence of candidate mechanisms. The inconsistent results may be attributed in part to the use of different mediation approaches across studies, which vary in power and required assumptions (Fritz & MacKinnon, 2007), compounded by the fact that RCTs are usually powered to detect between-group treatment effects but not necessarily mediation effects (Kraemer et al., 2002). Candidate processes other than therapeutic relationship variables have been omitted from systematic and meta-analytic reviews of EBPs for youth depression, though they were included in narrative reviews. In the past decade, which candidate processes and mechanisms have continued to be assessed, and how frequently, regardless of whether they were subjected to mediation tests? Are there more data on candidate mechanisms and processes for IPT? How can all these data gathered throughout the past several decades be harnessed and integrated to draw meaningful conclusions about how EBPs for youth depression work?

**The Present Systematic Review and Meta-Analysis**

We hope to address these questions with two complementary reviews: (1A) a systematic review of candidate change processes and mechanisms in RCTs of EBPs for youth depression, and (1B) a meta-analysis of selected candidates that have sufficient data, as mediators of treatment outcome. We term candidate change processes and mechanisms collectively as candidate mediators (CMs) because of our aims to examine them as mediators, but we use the terms therapy change processes, client change processes, and change mechanisms where their conceptual differences are relevant. Similar to previous work (e.g., Chu & Harrison, 2007; Weersing & Weisz, 2002), we focus on EBPs because they already have a solid foundation of
evidence for their efficacy, thus a logical next step would be to further our understanding of how they work to maximize their effects. We included CBT, IPT, and non-cognitive behavioral therapy (BT) as these therapies were designated as “well-established” or “probably efficacious” therapies for child or adolescent depression by the Society of Clinical Child and Adolescent Psychology, Division 53 of the American Psychological Association (David-Ferdon & Kaslow, 2008). Examining multiple EBPs could help shed light on CMs that are specific versus common to different theoretical orientations.

**Study 1A: Systematic Review**

The present systematic review involves a comprehensive search and description of CMs assessed in RCTs of EBPs for youth depression. We updated the two prior reviews by Weersing & Weisz (2002) and Chu and Harrison (2007) to include RCTs published up through 2013 and expand on the recent review by Weersing et al. (2016) to include CMs that were not subjected to mediation testing. We documented which CMs were assessed, how commonly they were assessed across RCTs, and methods of assessment (e.g., measures used, time points administered, whether tested as mediators, which mediation approaches). Here, we used descriptive statistics (e.g., frequencies, proportions) and qualitative examination of the information gathered to identify CMs that have, and have not, been adequately assessed, and to select CMs for the meta-analysis. We also noted whether researchers are testing mediation more frequently and using newer, improved mediation tests in the past decade since publication of several influential papers on psychotherapy mediators and mechanisms (e.g., Kazdin & Nock, 2008).

---

4 We followed David-Ferdon and Kaslow’s (2008) EBP designations because that was the most updated EBP review for youth depression when we started work on the present systematic review and meta-analysis. Moreover, these designations are consistent with those made in EBP reviews for adult depression (Craighead, Sheets, Brosse, & Ilardi, 2007; Strunk, n.d.). However, the Weersing et al. (2016) update designated CBT and IPT but not BT as an EBP after applying stricter inclusion criteria to RCTs (e.g., youth needed to meet diagnostic criteria, clinical cutoffs, or be identified by mental health professionals; RCT findings that failed to support the therapy’s efficacy counted against the designation of EBP status, whereas in previous reviews, findings that failed to support the therapy’s efficacy were ignored).
and discussed methodological recommendations for future research.

**Study 1A: Method**

**Inclusion and Exclusion Criteria**

To be included in the systematic review, RCTs had to meet the following criteria: (a) randomly assigned individual participants to treatment conditions; (b) tested CBT, IPT, or BT either as a standalone treatment, or in combination with another therapy (e.g., CBT plus separate parent group intervention); (c) employed a control condition consisting of waitlist/no treatment, attention or pill placebo, case management, or usual care; (d) sampled participants who were selected or referred due to diagnosis of depressive disorder or elevated symptoms of depression; (e) sampled children or adolescents with a mean participant age between 4 and 18 years; and (f) assessed at least one continuous outcome measure of depression symptoms. We included RCTs that met all the above criteria, regardless of whether they measured a CM or not so that those that measured a CM could be compared to those that did not in frequency and in treatment effects. Although the RCTs were limited to those with outcome papers published in English language peer-reviewed journals, once the RCTs were identified for inclusion, we included all materials containing methods or data from the RCTs in the present review. These materials included journal articles, dissertations, theses, and personal correspondence with RCT authors reporting study methodology, primary or follow-up outcomes, secondary analyses, and summary statistics, published or unpublished in any language, as well as deidentified datasets.

We excluded treatment conditions involving an EBP combined with medication because it is unclear how EBPs may interact with medication to improve symptoms and because it was unlikely that we would find a large number of studies assessing the same combination of EBP
plus medication that could be analyzed separately. Also excluded were RCTs that only compared an EBP to another EBP, to a psychotherapy with a specific theoretical rationale (beyond attention and support), or to a specific medication, as there would not be a clear control condition. In addition, we excluded RCTs that selected youths on the basis of exhibiting either depression symptoms or some other problem (e.g., anxiety, negative thinking) as the data from those studies may go beyond what is found with depressed youth, as well as RCTs that randomized only therapists or groups of youths, as the statistics yielded would not be comparable to that from RCTs that randomized individual youths to treatment conditions.

**Literature Search and Retrieval**

RCTs meeting the above criteria were drawn from our comprehensive meta-analytic database of RCTs testing psychological therapies for youth depression, anxiety, conduct problems, and attention-deficit hyperactivity disorder (ADHD), with outcome papers published from 1963 through 2013. Weisz et al. (2017) details the search strategies used to build this comprehensive database. Briefly, literature searches were conducted on the electronic databases PsycINFO and PubMed, using a set of 21 psychotherapy-related terms (e.g., psychother-, counseling) in PsycINFO, and the MeSH indexing system in PubMed, with search limits for clinical or outcome assessment, child and adolescent age group, human subjects, and published in English. The abstracts of every entry retrieved in the searches were reviewed by one research assistant, and the full text article was obtained if the abstract indicated a potentially eligible RCT. Then two research assistants reviewed each full-text article independently, resolving discrepancies through discussion or consultation with a third research assistant or graduate student. RCTs included in this database had to meet many of the inclusion criteria for the present review, including random assignment, child and adolescent participants, and appropriate control
condition. The subset of RCTs focused on depression, testing an EBP, and assessing a continuous measure of depression symptoms was selected based on coding by a single coder after establishing intercoder reliability with at least one other coder. The coding procedures and reliability statistics are reported in the next two sections.

Because the comprehensive database was focused on posttreatment and follow-up outcomes, they may have missed information on CMs reported in secondary analysis papers. Thus, we undertook additional strategies to locate relevant literature and data for the present systematic review and meta-analysis. For each included RCT, we used PsycINFO and PubMed to search for records of articles that cited the outcome paper(s), or those that were authored or co-authored by the first or second author of the outcome paper(s) from the year the earliest outcome paper was published. Duplicates were deleted, and abstracts were screened and used to decide whether to retrieve the full-text article, which were in turn reviewed to determine if they reported methods or data from any included RCT. We also screened the reference sections in articles from the included RCTs, and in reviews of youth depression EBPs, change processes, and change mechanisms, for any additional articles reporting information from the included RCTs. Finally, we contacted the authors of the included RCTs and the National Institute of Mental Health (NIMH) data repositories to request information (e.g., questions about measures and time points) and data (i.e., summary statistics or deidentified datasets) needed for coding of study characteristics and effect sizes. When authors did not respond or were unable to provide part or all the data, we retrieved dissertations or theses from the university libraries that store them, from Proquest Dissertations and Theses electronic database, or directly from the authors.
Figure 1.1 details the phases of identification, screening, and assessment of eligibility of RCTs, as well as relevant articles, dissertations, theses, and data obtained from RCT authors or data repositories, for inclusion in the present review.

**Coding of Study Characteristics**

Studies were coded for characteristics of participants, treatment/control conditions, outcome measures, and study quality as part of the comprehensive meta-analytic database (for detailed coding procedures, see Weisz et al., 2017). Coders were eight postdoctoral fellows and graduate students in psychology who each independently coded 20 to 30 RCTs to establish intercoder reliability before coding the remainder individually. Coder disagreements were resolved by discussion and consultation with one another. Mean kappa ($k$) was computed for categorical variables, and mean intraclass correlation (ICC) was computed for continuous variables. Among these coded variables, target problem ($k=.89$) and the treatment and control condition codes ($k=.83$) were used to facilitate the selection of RCTs that targeted depression and that tested CBT, IPT, or BT against an appropriate control condition for inclusion into the present review. We used other coded variables to compute descriptive statistics of included RCT participants: mean participant age (ICC=.99), youngest age (ICC=1.00), oldest age (ICC=.99), percent of each gender (ICC=.95), percent participants who are Caucasian (ICC=.87), African American (ICC=1.00), Latino (ICC=.86), Asian (ICC=.70), Native American (ICC=.57), and of other (ICC=.73) and unknown race/ethnicity (ICC=.84), geographical location ($k=.88$), recruited vs. clinically referred/treatment-seeking vs. court-mandated or incarcerated ($k=.66$), and all participants required vs. not required to have a diagnosis of depression ($k=.79$).
Figure 1.1. Flowchart detailing the identification, screening, assessment for eligibility of randomized controlled trials (RCTs), as well as relevant articles, dissertations, theses, and data obtained from RCT authors or data repositories, for inclusion in the present review.
4,017 Full-text primary outcome papers retrieved and assessed for comprehensive meta-analytic database

8,926 Records identified through electronic database searches for each included RCT

3,332 Failed to meet inclusion criteria for meta-analytic database (no random assignment, participants not youths with diagnosis or elevated problems, target problem not depression, anxiety, conduct problems, or attention-deficit/hyperactivity, no psychological therapy condition, failed multiple inclusion criteria)

685 RCTs eligible for inclusion in meta-analytic database

3,887 Removed (duplicate record, authors do not include first or second author of outcome/follow-up papers)

5,039 Abstracts screened

625 Failed to meet inclusion criteria for present systematic review and meta-analysis (no appropriate control condition, target problem not depression, target problem depression or other disorder, no evidence-based psychotherapy for youth depression)

67 Articles from the 40 included RCTs previously identified in comprehensive meta-analytic database

226 Full-text articles retrieved and assessed whether reporting methods/results from included RCTs

40 RCTs included in present systematic review and meta-analysis

168 Not reporting methods/results from included RCT (e.g., cites non-included RCT as data source, discrepant treatment conditions or participant characteristics)

4,017 Full-text primary outcome papers retrieved and assessed for comprehensive meta-analytic database

8,926 Records identified through electronic database searches for each included RCT

3,332 Failed to meet inclusion criteria for meta-analytic database (no random assignment, participants not youths with diagnosis or elevated problems, target problem not depression, anxiety, conduct problems, or attention-deficit/hyperactivity, no psychological therapy condition, failed multiple inclusion criteria)

685 RCTs eligible for inclusion in meta-analytic database

3,887 Removed (duplicate record, authors do not include first or second author of outcome/follow-up papers)

5,039 Abstracts screened

625 Failed to meet inclusion criteria for present systematic review and meta-analysis (no appropriate control condition, target problem not depression, target problem depression or other disorder, no evidence-based psychotherapy for youth depression)

67 Articles from the 40 included RCTs previously identified in comprehensive meta-analytic database

226 Full-text articles retrieved and assessed whether reporting methods/results from included RCTs

168 Not reporting methods/results from included RCT (e.g., cites non-included RCT as data source, discrepant treatment conditions or participant characteristics)
9 Dissertations or theses reporting method/results from included RCTs retrieved from databases or libraries, or provided by RCT authors

57 Articles reporting method/results from included RCTs identified through RCT-specific searches

24 Sets of summary statistics or deidentified datasets obtained through personal correspondence with RCT authors or data repositories

3 Articles reporting method/results from included RCTs identified through other sources (e.g., reference trails, colleagues)

4 Articles repeated (i.e., each reporting method/results from two different included RCTs)

123 Articles, 9 dissertations/theses, and 24 sets of data included in present systematic review and meta-analysis
Identification and Coding of Candidate Processes and Mechanisms

Additional screening and coding were performed to identify and characterize CMs and outcome measures by undergraduate or post-Bachelor’s degree research assistants. All the literature retrieved from the included RCTs were screened to determine whether any formal tests of mediation were conducted, in which the predictor is treatment condition, the mediator variable is a CM, and the outcome is a depression symptom outcome measure. The four research assistants who screened the articles for tests of treatment mediation achieved substantial agreement on the 20 RCTs they each coded ($k=.75$).

Next, CMs were identified, regardless of whether they were tested as mediators by RCT authors, together with suitable outcome measures, for further coding. Selection and coding of change processes and mechanisms, and outcomes, were performed by three research assistants, who each coded 2 RCTs for practice, and another 12 RCTs for establishing intercoder reliability, before individually coding the rest of the sample. In order to cast a wide net for CMs, coders were instructed to review all materials from an RCT and note all variables measured for both the treatment and control group that could potentially change as a result of treatment (regardless of whether there was change, or whether the variable was measured more than once). We excluded the following variables: (a) participant demographics, side effects or adverse events, service use, economic and time cost measures, parent or sibling variables that do not pertain to the youth or to their relationship with the youth, outcome measures of overall symptoms, of a target problem other than depression, or of general youth functioning because these are not CMs or outcomes of interest; (b) percent of participants meeting diagnosis or clinical cutoff for a depressive disorder because our mediation model includes only continuous mediator and outcome variables; (c) time-to-event variables (e.g., length of time to recurrence of disorder) because they are not

---

5 Mean percentage agreement was 93.3%.
compatible with the designated time points in hypothesized mediation model; (d) measures that are designed for a specific treatment and thus administered to only one treatment group, because mediation effects in RCTs are examined for one treatment in relation to the other (instead of within one treatment condition), and mediation analysis requires the mediator variable to vary across treatment conditions (Kraemer et al., 2002); (e) treatment session attendance because of its unclear meaning—it could reflect dose of therapy, youth or parent engagement, or the therapists’ judgment of the amount of treatment needed for a particular youth if number of sessions is not fixed—and it could mean different things across different RCTs; and (f) treatment satisfaction, as this is more like a non-symptom outcome than a therapy change process or mechanism. The retained variables were then selected for further coding if they were continuous measures of CMs or of depression symptoms including suicidality (i.e., outcomes); internalizing symptom measures were selected only if purely depression symptom measures were unavailable. Intercoder agreement on variable selection was substantial ($k=.69$).

We created 33 possible mediator categories, of which 21 were used ($k=91$; see Table 1.1). These mediator categories were created with reference to reviews of change processes and mechanisms in therapies for youth depression (e.g., Chu & Harrison, 2007; Weersing & Weisz 2002) and adult depression (e.g., Crits-Christoph et al., 2013; Gu et al., 2015), and to chapters describing EBPs for youth depression (e.g., Clarke & DeBar; 2010, Jacobson & Mufson, 2010) and adult depression (e.g., Beach, Jones, & Franklin, 2009; Dimidjian, Martell, Addis, & Herman-Dunn, 2008). We also included an “Other” category to capture plausible change processes and mechanisms that did not fit into existing categories, and then subsequently decided by consensus whether to combine each one with an existing mediator category or create a new category for it. We tended towards being more inclusive in our mediator category codes due to
research indicating that EBPs can have effects on CMs not theoretically targeted by that EBP (e.g., Weersing et al., 2009).

Table 1.1 *Candidate Mediators (CM) and the Frequency with Which They were Measured at Various Time Points in the Included Sample of 40 Randomized Controlled Trials*

<table>
<thead>
<tr>
<th>Mediator Category</th>
<th>Explanation &amp; Examples</th>
<th>Putative Role</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Overall Pre Mid Post FU</td>
<td></td>
</tr>
<tr>
<td><strong>Negative Thinking</strong></td>
<td>Cognitive distortions, dysfunctional attitudes, pessimistic attributional style, external locus of control, negative info-processing bias, hopelessness, rumination, negative self-concept/self-esteem, perfectionism.</td>
<td>Change mechanism of CBT</td>
<td>24 24 4 24 14</td>
</tr>
<tr>
<td><strong>Social Skills and Relations</strong></td>
<td>Participation in social activities, positive self-presentation, social adjustment, interpersonal relationship quality, social support, loneliness, sociometric ratings of liking, prosocial behavior.</td>
<td>Change mechanism of CBT, IPT, and BT</td>
<td>14 14 3 13 6</td>
</tr>
<tr>
<td><strong>Family Dysfunction</strong></td>
<td>Family/parent expressed emotion, communication problems, relationship quality, lack of cohesion.</td>
<td>Change mechanism of IPT</td>
<td>13 13 2 12 8</td>
</tr>
<tr>
<td><strong>Problem Solving</strong></td>
<td>Skill or effort taken to change external conditions, active or primary control coping, conflict resolution skills, generating solutions, asking others for help to solve a problem, family problem solving, action stage of change.</td>
<td>Change mechanism of CBT, IPT, and BT</td>
<td>8 8 3 8 4</td>
</tr>
<tr>
<td><strong>Treatment Expectancy</strong></td>
<td>Extent to which one expects to benefit from treatment or prefers a treatment over another, attitudes towards treatment.</td>
<td>Client change process and common factor</td>
<td>7 6 2 1 0</td>
</tr>
<tr>
<td>Mediator Category</td>
<td>Explanation &amp; Examples</td>
<td>Putative Role</td>
<td>Frequency</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Overall</td>
<td>Pre</td>
</tr>
<tr>
<td>Cognitive Skills</td>
<td>Skill or effort in changing thinking to deal with stressors or feelings, secondary control coping, distraction, finding the silver lining, alternative explanations, acceptance, identifying negative thoughts.</td>
<td>Change mechanism of CBT</td>
<td>5 5 2 4 1</td>
</tr>
<tr>
<td>Pleasant Activities</td>
<td>Activity scheduling, engagement in enjoyable activities in general (though not specifically social activities).</td>
<td>Change mechanism of CBT and BT</td>
<td>5 5 1 5 4</td>
</tr>
<tr>
<td>Avoidant Coping</td>
<td>Wishful thinking, denial or avoidance of the problem, precontemplation stage of change.</td>
<td>Change mechanism of CBT</td>
<td>4 4 2 4 1</td>
</tr>
<tr>
<td>Role Engagement</td>
<td>Youth or parent involvement in therapy, treatment compliance, homework completion, knowledge of therapy content.</td>
<td>Client change process and common factor</td>
<td>3 1 1 2 0</td>
</tr>
<tr>
<td>Impulsive Coping</td>
<td>Finding solutions to problems in a hurried, unsystematic manner, inadequate outcomes monitoring outcomes.</td>
<td>Change mechanism not targeted by depression EBPs</td>
<td>2 2 2 2 0</td>
</tr>
<tr>
<td>Therapeutic Alliance</td>
<td>The collaborative and affective relationship between the therapist and the parent/youth, agreement between therapist and parent/youth about goals, techniques implemented to achieve the goals, and the therapist–parent/youth bond.</td>
<td>Therapy–client change process and common factor</td>
<td>2 1 1 1 0</td>
</tr>
<tr>
<td>Motivation to Change</td>
<td>Youth desire to reduce depression or change behavior, contemplation stage of change.</td>
<td>Candidate mechanism and common factor</td>
<td>2 2 1 0 0</td>
</tr>
<tr>
<td>Expression of Affect</td>
<td>Talking/writing about feelings, communicating it through other creative outlets.</td>
<td>Change mechanism of IPT</td>
<td>1 1 1 1 0</td>
</tr>
</tbody>
</table>
Table 1.1 (Continued).

<table>
<thead>
<tr>
<th>Mediator Category</th>
<th>Explanation &amp; Examples</th>
<th>Putative Role</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Overall</td>
<td>Pre</td>
</tr>
<tr>
<td>Maintenance of Gains</td>
<td>Youth effort to consolidate behavioral change, maintenance stage of change.</td>
<td>Candidate mechanism and common factor</td>
<td>1 1 1 0 0</td>
</tr>
<tr>
<td>Therapist CBT Strategies</td>
<td>Therapist adherence to, competence in, or knowledge of a CBT protocol.</td>
<td>Therapy change process of CBT</td>
<td>1 0 1 0 0</td>
</tr>
<tr>
<td>Therapist Client-Centered Strategies</td>
<td>Therapist adherence to, competence in, or knowledge of a psychodynamic therapy protocol.</td>
<td>Therapy change process of non-EBP</td>
<td>1 0 1 0 0</td>
</tr>
<tr>
<td>Therapist Family Systems Strategies</td>
<td>Therapist adherence to, competence in, or knowledge of a family systems therapy protocol</td>
<td>Therapy change process of non-EBP</td>
<td>1 0 1 0 0</td>
</tr>
<tr>
<td>Therapist Psychodynamic Strategies</td>
<td>Therapist adherence to, competence in, or knowledge of a psychodynamic therapy protocol.</td>
<td>Therapy change process of non-EBP</td>
<td>1 0 1 0 0</td>
</tr>
<tr>
<td>Group Cohesiveness</td>
<td>Youth perception of belonging to the group, acceptance by group members, and attractiveness of the group.</td>
<td>Client change process and common factor</td>
<td>1 0 1 0 0</td>
</tr>
<tr>
<td>Research Understanding</td>
<td>Youth and parent knowledge of RCT purpose, design, treatment options, consent process, benefits and risks.</td>
<td>Change mechanism not targeted by depression EBPs</td>
<td>1 0 1 0 0</td>
</tr>
<tr>
<td>Perceived Stigma</td>
<td>Youth perception of acceptance by others if they thought the youth had depression.</td>
<td>Change mechanism not targeted by depression EBPs</td>
<td>1 1 0 0 0</td>
</tr>
</tbody>
</table>
Table 1.1 (Continued).

Note. CBT = cognitive behavioral therapy, IPT = interpersonal therapy, BT = behavioral therapy, pre = pretreatment assessment, mid = during treatment assessment, post = posttreatment assessment, FU = follow-up assessment.

All selected outcomes and CMs were also coded for whether their source was the youths themselves, caregivers, siblings, peers, teachers, therapists, researchers, or life-event data (k=.89); and whether they were scaled in the default direction of higher scores reflecting higher levels of the construct (e.g., higher scores by default reflects more severe symptomatology, more frequent negative thinking, more socially skilled) (k=.82). Coders also indicated whether the outcomes and CMs were measured at pretreatment (k =.94), during treatment (k =.88; which we term the “midtreatment” time point though the measure could have been given at any time during treatment), at posttreatment (k =.72), and at follow-up (k=.73), and number of weeks from pretreatment to each subsequent time point (ICC=.86–.98). If multiple follow-up time points were available, then coders were instructed to select the one closest to six months after posttreatment assessment because this is the most common follow-up time point among depression RCTs in the comprehensive meta-analytic database (44% of RCTs with measures given at any follow-up time point had a follow-up time point at six months posttreatment).

Study 1A: Results

Description of Study Sample

Forty RCTs, testing a total of 5,608 youths in 50 EBP conditions and 44 eligible control conditions, met inclusion criteria (see Table 1.2). The searches for literature relating to these
included RCTs yielded a total of 123 journal articles published from 1986 through 2014. The authors of all but one RCT (Reed, 1994) responded to requests for data and information, of whom 30 contributed data, provided dissertations/theses, or directed us to supplementary materials. We obtained unpublished summary statistics for 17 RCTs, deidentified datasets for 7 RCTs, 7 doctoral dissertations (including one, Tang, 2006 published in Chinese), and 2 masters/honors theses to supplement the information from published journal articles.

The pooled participant sample had a mean age of 14.0 years (SD=2.2); the youngest and oldest participants across all RCTs were 3 and 22 years old respectively. On average, under half the youths were male (M=39.9%, SD=12.9). Caucasian youths made up less than half the sample (M=43.2%, SD=37.8), with other youths identifying as Latino (M=14.1%, SD=26.7), African American (M=7.6% , SD=12.9), Asian (M=3.7% , SD=9.2), Native American (M=0.3%, SD=1.6), other race/ethnicity (M=12.2%, SD=24.7), or of unknown race/ethnicity (M=19.0%, SD=38.0). The majority (75.0%) of RCTs was conducted in North America, notable minorities in Europe (10.0%) and in Australia or New Zealand (7.5%), two in Asia (5.0%), and one in Africa (2.5%); none were conducted in South America. Most (72.5%) participant samples comprised youths who were recruited for the RCT but otherwise not seeking or receiving mental health services, 20.0% comprised clinically referred or treatment-seeking youths, 2.5% comprised court-mandated or incarcerated youths, and 5.0% comprised youths recruited through unknown means. A third (30.0%) of the RCTs required youths to have a diagnosis of depression, the remainder did not require a depression diagnosis, required an absence of depression diagnosis, or did not report on diagnostic requirements.

More than three-quarters (78.0%) of treatment conditions involved CBT, with IPT (12.0%) and BT (10.0%) tested much less frequently. Two-thirds of the 39 CBT conditions were
delivered in group format; the rest were delivered individually (17.9%), as bibliotherapy (7.7%), as a computerized program (5.1%), or as a single family-based session (2.6%). The 5 BT conditions were also mostly delivered as group interventions, focusing either on relaxation (40%), on social skills training (20%), or on activity scheduling and problem solving (20%). There was one individual BT condition (20%)—“self-modeling,” in which youths watched themselves on edited videotapes displaying desired target behaviors (e.g., smiling, verbalizing positive self-attributions). In contrast to CBT and BT, the six IPT conditions were evenly divided into group and individual format. The most common control condition was no treatment or waitlist (38.6%), followed by usual care (36.4%), and attention placebo (20.5%); pill placebo (2.3%), and case management (2.3%) were each employed in only one control condition.
Table 1.2 Study Sample, Treatment Conditions, and Outcomes, Candidate Mediator Categories, and Assessment Time Points

<table>
<thead>
<tr>
<th>Outcome Paper</th>
<th>N</th>
<th>EBP</th>
<th>Control</th>
<th>Outcome</th>
<th>Neg Think</th>
<th>Social</th>
<th>Family</th>
<th>Prob Solv</th>
<th>Cog Skills</th>
<th>Pleas Act</th>
<th>Avoid</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ackerson, Scogin, McKendree-Smith &amp; Lyman (1998)</td>
<td>22</td>
<td>CBT bibliotherapy</td>
<td>Waitlist</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Role Engagement</td>
</tr>
<tr>
<td>Asarnow et al. (2005)</td>
<td>418</td>
<td>CBT individual (enhanced collaborative primary care)</td>
<td>Usual care (in primary care + brief provider training)</td>
<td>Pre, post, FU</td>
<td>Pre</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Treatment Expectancy, Perceived Stigma</td>
</tr>
<tr>
<td>Asarnow et al. (2011)</td>
<td>181</td>
<td>CBT family-based (emergency care linkage to outpatient treatment)</td>
<td>Usual care (in emergency department + brief provider training)</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asarnow, Scott, &amp; Mintz (2002)</td>
<td>23</td>
<td>CBT group</td>
<td>Waitlist</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bolton et al. (2007)</td>
<td>209</td>
<td>IPT group</td>
<td>Waitlist</td>
<td>Pre, post</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brent et al. (1997)</td>
<td>72</td>
<td>CBT individual</td>
<td>Attention placebo</td>
<td>Pre, mid, post, FU</td>
<td>Pre, mid, post</td>
<td>Pre, mid, post</td>
<td>Pre, mid, post</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Treatment Expectancy, Role Engagement</td>
</tr>
<tr>
<td>Clarke et al. (1995)</td>
<td>150</td>
<td>CBT group + usual care</td>
<td>Usual care</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarke et al. (2001)</td>
<td>94</td>
<td>CBT group + usual care</td>
<td>Usual care (in HMO)</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarke et al. (2002)</td>
<td>88</td>
<td>CBT group + usual care</td>
<td>Usual care (in HMO)</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.2 (Continued).

<table>
<thead>
<tr>
<th>Outcome Paper</th>
<th>N</th>
<th>EBP</th>
<th>Control</th>
<th>Outcome</th>
<th>Neg Think</th>
<th>Social</th>
<th>Family</th>
<th>Prob Solv</th>
<th>Cog Skills</th>
<th>Pleas Act</th>
<th>Avoid</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarke, Rohde, Lewinsohn, Hops, &amp; Seeley (1999)</td>
<td>123</td>
<td>1. CBT group 2. CBT group + parent group</td>
<td>Waitlist</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fleming, Dixon, Frampton &amp; Merry (2012)</td>
<td>30</td>
<td>CBT computerized</td>
<td>Waitlist</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garber et al. (2009)</td>
<td>316</td>
<td>CBT group + usual care</td>
<td>Usual care</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gillham et al. (2007)</td>
<td>697</td>
<td>CBT group</td>
<td>1. Attention placebo 2. No treatment</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gillham et al. (2012)</td>
<td>408</td>
<td>1. CBT group 2. CBT group + parent group</td>
<td>Usual care</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome Paper</td>
<td>N</td>
<td>EBP</td>
<td>Control</td>
<td>Outcome</td>
<td>Neg Think</td>
<td>Social</td>
<td>Family</td>
<td>Prob Solv</td>
<td>Cog Skills</td>
<td>Pleas Act</td>
<td>Avoid</td>
<td>Other</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------</td>
<td>------</td>
<td>--------------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>-----------</td>
<td>---------</td>
<td>---------</td>
<td>-----------</td>
<td>------------</td>
<td>-----------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Gillham, Hamilton, Freres, Patton, &amp; Gallop (2006)</td>
<td>171</td>
<td>CBT group</td>
<td>Usual care</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green et al. (2011)</td>
<td>366</td>
<td>CBT group + usual care</td>
<td>Usual care</td>
<td>Pre, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kahn, Kehle, Jenson, &amp; Clark (1990)</td>
<td>68</td>
<td>1. BT group (relaxation)</td>
<td>Waitlist</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. BT group (self-modeling)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. CBT group + parent group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lewinsohn, Clarke, Hops, &amp; Andrews (1990)</td>
<td>74</td>
<td>1. CBT group</td>
<td>Waitlist</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. CBT group + parent group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liddle &amp; Spence (1990)</td>
<td>31</td>
<td>CBT group</td>
<td>1. Waitlist 2. Attention placebo</td>
<td>Pre, post, FU</td>
<td>Pre, post</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McCarty, Violette, Duong, Cruz, &amp; McCauley (2013)</td>
<td>120</td>
<td>CBT group</td>
<td>Attention placebo</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Merry et al. (2012)</td>
<td>30</td>
<td>computerized</td>
<td>Usual care</td>
<td>Pre, post, FU</td>
<td>Pre, post</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.2 (Continued).

<table>
<thead>
<tr>
<th>Outcome Paper</th>
<th>N</th>
<th>EBP</th>
<th>Control</th>
<th>Outcome</th>
<th>Neg Think</th>
<th>Social</th>
<th>Family</th>
<th>Prob Solv</th>
<th>Cog Skills</th>
<th>Pleas Act</th>
<th>Avoid</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mufson, Myrna, Weissman, Moreau, &amp; Garfinkel, (1999)</td>
<td>48</td>
<td>IPT individual</td>
<td>Attention placebo</td>
<td>Pre, mid, post</td>
<td>Pre, mid, post</td>
<td>Pre, mid, post</td>
<td>Pre, mid, post</td>
<td>Pre, mid, post</td>
<td>Pre, mid, post</td>
<td>Impulsive Coping</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reed (1994)</td>
<td>18</td>
<td>BT group (social skills training)</td>
<td>Attention placebo</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reynolds &amp; Coats (1986)</td>
<td>30</td>
<td>1. BT group (relaxation) 2. CBT group</td>
<td>Waitlist</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rohde, Clarke, Mace, Jorgensen, &amp; Seeley (2004)</td>
<td>93</td>
<td>CBT group</td>
<td>Case management</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Group Cohesiveness, Therapeutic Alliance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rosselló &amp; Bernal (1999)</td>
<td>71</td>
<td>1. CBT individual 2. IPT individual</td>
<td>Waitlist</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stark, Reynolds, &amp; Kaslow (1987)</td>
<td>29</td>
<td>1. BT group (behavioral problem-solving training) 2. CBT group</td>
<td>Waitlist</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Pre, post</td>
<td>Treatment Expectancy</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.2 (Continued).

<table>
<thead>
<tr>
<th>Outcome Paper</th>
<th>N</th>
<th>EBP</th>
<th>Control</th>
<th>Outcome</th>
<th>Neg Think</th>
<th>Social</th>
<th>Family</th>
<th>Prob Solv</th>
<th>Cog Skills</th>
<th>Pleas Act</th>
<th>Avoid</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stice, Rohde, Seeley, &amp; Gau (2008)</td>
<td>253</td>
<td>1. CBT group</td>
<td>No treatment (education brochure)</td>
<td>Pre, mid, post, FU</td>
<td>Pre, mid, post, FU</td>
<td>Pre, mid, post, FU</td>
<td>Pre, mid, post, FU</td>
<td>Pre, mid, post, FU</td>
<td>Expression of Affect, Motivation to Change</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Szigethy et al. (2007)</td>
<td>41</td>
<td>CBT individual</td>
<td>Waitlist</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tang, Jou, Ko, Huang, &amp; Yen (2009)</td>
<td>73</td>
<td>IPT individual</td>
<td>Usual care (school counseling)</td>
<td>Pre, post</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment for Adolescents With Depression Study Team (2004)</td>
<td>223</td>
<td>CBT individual</td>
<td>Pill placebo</td>
<td>Pre, mid, post</td>
<td>Pre, mid, post</td>
<td>Pre, mid, post</td>
<td>Pre, mid, post</td>
<td>Pre, mid, post</td>
<td>Treatment Expectancy, Impulsive Coping, Motivation to Change, Maintenance of Gains, Research Understanding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vostanis, Feehan, Grattan, &amp; Bickerton (1996b)</td>
<td>57</td>
<td>CBT individual</td>
<td>Attention placebo</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Treatment Expectancy, Role Engagement</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Attention placebo (journaling)
Table 1.2 (Continued).

<table>
<thead>
<tr>
<th>Outcome Paper</th>
<th>N</th>
<th>EBP</th>
<th>Control</th>
<th>Outcome</th>
<th>Neg Think</th>
<th>Social</th>
<th>Family</th>
<th>Prob Solv</th>
<th>Cog Skills</th>
<th>Pleas Act</th>
<th>Avoid</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weisz et al. (2009)</td>
<td>57</td>
<td>CBT individual</td>
<td>Usual care</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td>Pre, post, FU</td>
<td></td>
<td>Treatment Expectancy, Therapeutic Alliance, Therapist Strategies - CBT, Client-centered, Family Systems, and Psychodynamic</td>
</tr>
<tr>
<td>Weisz, Thurber, Sweeney, Proffitt, &amp; LeGagnoux (1997)</td>
<td>48</td>
<td>CBT group</td>
<td>No treatment</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Treatment Expectancy</td>
</tr>
<tr>
<td>Wood, Trainor, Rothwell, Moore, &amp; Harrington (2001)</td>
<td>63</td>
<td>CBT group + usual care</td>
<td>Usual care</td>
<td>Pre, mid, post</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young, Mufson, &amp; Davies (2006)</td>
<td>41</td>
<td>IPT group</td>
<td>Usual care (school counseling)</td>
<td>Pre, mid, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young, Mufson &amp; Gallop (2010)</td>
<td>57</td>
<td>IPT group</td>
<td>Usual care (school counseling)</td>
<td>Pre, mid, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yu &amp; Seligman (2002)</td>
<td>220</td>
<td>CBT group</td>
<td>Waitlist</td>
<td>Pre, post, FU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.2 (Continued).

*Note.* EBP = evidence-based psychotherapy, Neg Think = negative thinking, Social = social skills and relations, Family = family dysfunction, Prob Solv = problem solving, Cog Skills = cognitive skills, Pleas Act = pleasant activities, Avoid = avoidant coping, CBT = cognitive behavioral therapy, IPT = interpersonal therapy, BT = behavioral therapy, pre = pretreatment assessment, mid = during treatment assessment, post = posttreatment assessment, FU = follow-up assessment, HMO = health maintenance organization.
Description of Outcomes and Candidate Mediators

We coded a total of 92 outcome measures from the 40 RCTs, and 212 CMs from 30 (75.0%) of the RCTs, administered to participants at any time point. All 40 RCTs measured at least one outcome at pretreatment, 39 (97.5%) measured at least one outcome at posttreatment, 27 (67.5%) measured at least one outcome at follow-up, and only 7 (17.5%) measured at least one outcome during treatment. The relative frequency with which CMs were measured at each time point mirrored those of outcomes: all 30 RCTs that measured any CM measured at least one CM at pretreatment, of which 28 (93.3%) measured at least one CM at posttreatment, 18 (60.0%) measured at least one CM at follow-up, and only 8 (26.7%) measured at least one CM during treatment. On average, the midtreatment, posttreatment, and follow-up measures were given at 5.7 weeks (SD=4.0), 12.2 weeks (SD=8.1), and 36.0 weeks (SD=17.7) after pretreatment. The most common informant for both outcomes and CM was the youth (outcome 70.7%, CM 75.9%), then the parent or caregiver (outcome 15.2%, CM 17.5%), research staff (outcome 13.0%, CM 4.2%), and school staff (outcome 1.1%, CM 0.9%). Small numbers of CMs, but no outcomes, were also completed by therapists (0.9%), and an unreported informant (0.5%).

The 212 CMs fall under 21 of 33 possible mediator categories, with descriptions and frequency of measurement at various time points listed in Table 1.1. Negative thinking is by far the most common mediator category, with 24 (60.0%) RCTs measuring it at posttreatment, making it nearly twice as frequent than the second and third most common categories—social skills and relations, and family dysfunction, measured by 14 RCTs (35.0%) and 13 RCTs (32.5%), respectively. Next are problem solving (8 RCTs, 20.0%), treatment expectancy (7 RCTs, 17.5%), cognitive skills and pleasant activities (5 RCTs each, 12.5%), avoidant coping (4 RCTs, 10.0%), and role engagement (3 RCTs, 7.5%). The least common mediator categories are
impulsive coping, therapeutic alliance, and motivation to change, each measured by 2 RCTs (5.0%); and expression of affect, maintenance of gains, therapist CBT, psychodynamic, family systems, and client-centered strategies, group cohesiveness, research understanding, and perceived stigma, each measured by 1 RCT (2.5%).

In Table 1.1, we labeled these 21 mediator categories as putative therapy change processes, putative client change processes, or putative change mechanisms according to Doss’ (2004) model. We also indicated whether each mediator category was considered by researchers to be specifically activated by CBT, IPT, BT, or non-EBPs, a “common factor” shared across interventions, or not targeted by depression EBPs, with reference to the sources used to generate the mediator categories described earlier (e.g., Clarke & DeBar; 2010, Crits-Christoph et al., 2013). Of the 21 mediator categories, 13 were putative change mechanisms, including four specific to CBT or BT (e.g., negative thinking, pleasant activities), two specific to IPT (e.g., family dysfunction), two for all three EBPs (e.g., social skills and relations), two common factors (e.g., motivation to change), and three not typically targeted by the EBPs (e.g., perceived stigma). Three categories were putative client change processes that were also common factors (e.g., role engagement), three categories were putative therapy change processes of non-EBPs (e.g., therapist client-centered strategies), one category was a putative therapy change process of CBT (i.e., therapist CBT strategies), and one category was a putative therapy–client change process and common factor (i.e., therapeutic alliance).

The relative frequencies of mediator categories at both pretreatment and posttreatment largely parallel their overall relative frequency in RCTs; only treatment credibility was assessed much more frequently at pretreatment (6 RCTs) than at posttreatment (1 RCT). Although midtreatment assessments were conducted for all but one mediator category (perceived stigma),
only a handful of RCTs (1 to 4 per category) conducted them. Follow-up assessments were conducted for all but one change mechanism mediator category (expression of affect). Understandably, follow-up assessments are unavailable for change process mediator categories because change processes, by definition, occur during treatment.

The other 12 mediator categories were not measured in any of the RCTs. These include hypothesized change mechanisms of IPT (i.e., insight into self and interpersonal patterns, understanding and finding meaning in loss and change) and hypothesized change mechanisms of non-EBPs (i.e., autobiographic memory specificity, mindfulness, self-compassion, psychological flexibility, cognitive and emotional reactivity, meta-awareness and decentering). Hypothesized therapy change processes that are specific to IPT (i.e., therapist IPT strategies) or that are nonspecific common factors (i.e., therapist role engagement, therapist flexibility, therapist directiveness) were also not measured in any RCT.

**Mediation Assessed by Study Authors**

In spite of the fact that 75% of the RCT sample measured a CM at any time point, with 70% measuring a CM at mid- or posttreatment, only 15% (6 RCTs) assessed any CM in formal mediation analysis. Although mediation testing is rare in RCTs of EBPs for youth depression, it appears to have increased slightly in the decade since the publication of review and conceptual papers calling for further study of treatment mediators and change mechanisms (e.g., Weersing & Weisz, 2002). From the 1980s through 2002, only three articles (Ackerson et al., 1998; Kolko, Brent, Baugher, Bridge, & Birmaher, 2000; Yu & Seligman, 2002) reported testing for treatment mediators in three RCTs, of which two tested only measures of negative thinking. Thereafter, five more articles (Dietz et al., 2014; Jacobs et al., 2009; Kaufman Rohde, Seeley, Clarke, & Stice, 2005; Lewis et al., 2009; Stice, Rohde, Seeley, & Gau, 2010) reporting assessments of
treatment mediators from four RCTs have been published, with CMs assessed coming from multiple categories.

As expected, the three early mediation articles published through 2002 all used Baron and Kenny’s (1986) causal steps approach. Among the five later mediation articles, two (Jacobs et al., 2009; Lewis et al., 2009) used another causal steps approach advocated by Kraemer et al. (2002), which builds on the Baron and Kenny approach by adding logical criteria (e.g., mediator should be a postrandomization construct) and a step testing for moderated mediation (i.e., if the mediator variable or the mediator variable × treatment condition interaction predicts outcome, then the criterion of the mediator predicting the outcome would be met). The other three used some combination of a causal steps approach plus a significance test of the indirect effect. Specifically, Kaufman et al. (2005) used Baron and Kenny’s approach; when criteria were met in the earlier steps, then the authors tested whether the difference between the total effect and direct effect (equivalent to the indirect effect) is different from zero. Stice et al. (2010) modified Baron and Kenny’s approach such that change in the mediator variable needs to predicts change in the outcome in the intervention condition only, rather than in the full sample; and when criteria were met in earlier steps, tested the difference between the total and direct effects; and then when the difference is significant, they assessed whether meaningful change in the mediator variable preceded meaningful change in the outcome (see Stice, Presnell, Gau, & Shaw, 2007). Finally, Dietz et al. (2014) employed a causal steps approach similar to Baron and Kenny’s, testing the significance of the indirect effect using bias-correct bootstrapping to generate confidence intervals when earlier steps were met.
The findings documented in these eight articles are summarized in Table 1.3. One of the RCTs (Stice et al., 2008) tested two eligible treatment conditions (CBT group and CBT bibliotherapy), thus mediation findings were reported separately for each treatment–control pair.

The most commonly assessed mediator category is negative thinking, mirroring its high frequency of measurement among RCTs. All six RCTs assessed negative thinking as a mediator, testing a total of 12 measures, of which 5 were found to be significant treatment mediators. The findings are mixed not only between studies, but also within studies, and even within specific measures. For example, Kaufman et al. (2005) and Stice et al. (2010) found that negative automatic thoughts mediated reductions in depression symptoms when comparing CBT group to control, whereas Ackerson et al. (1998) and Stice et al. (2010) found that they did not when comparing CBT bibliotherapy to control. However, treatment delivery format does not seem to explain discrepancy as Ackerson et al. (1998) found that dysfunctional attitudes was a mediator for CBT bibliotherapy (though only for one of three depression outcome measures), whereas Kaufman et al. (2005) and Jacobs et al. (2009) found that they were not a mediator for CBT group and CBT individual respectively.
<table>
<thead>
<tr>
<th>Outcome Paper</th>
<th>Treatment Conditions</th>
<th>Treatment Effect?</th>
<th>Mediation Paper</th>
<th>Candidate Mediator</th>
<th>Significant Mediator?</th>
<th>Finding (treatment effects described reflect comparison to control)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Negative Thinking</strong></td>
<td>Ackerson et al. (1998)</td>
<td>Yes</td>
<td>Ackerson et al. (1998)</td>
<td>Automatic Thoughts Questionnaire (ATQ)</td>
<td>No</td>
<td>CBT did not reduce negative thoughts</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dysfunctional Attitude Scale (DAS)</td>
<td>Yes</td>
<td>CBT reduced dysfunctional attitudes, which mediated decrease in depression symptoms measured by the CDI but not CBCL-D or HRSD</td>
</tr>
<tr>
<td></td>
<td>Brent et al. (1997)</td>
<td>Yes</td>
<td>Kolko et al. (2000)</td>
<td>Children's Negative Cognitive Error Questionnaire (CNCEQ)</td>
<td>No</td>
<td>CBT reduced cognitive distortions, but authors did not continue mediation test as treatment effect was nonsignificant in mediation paper</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Beck Hopelessness Scale (BHS)</td>
<td>No</td>
<td>CBT did not reduce hopelessness</td>
</tr>
<tr>
<td></td>
<td>Rohde et al. (2004)</td>
<td>Yes</td>
<td>Kaufman et al. (2005)</td>
<td>ATQ</td>
<td>Yes</td>
<td>CBT reduced negative thoughts, which mediated decrease in depression symptoms</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DAS (brief)</td>
<td>No</td>
<td>CBT did not reduce dysfunctional attitudes</td>
</tr>
<tr>
<td></td>
<td>Stice et al. (2008)</td>
<td>Yes</td>
<td>Stice et al. (2010)</td>
<td>ATQ (brief)</td>
<td>Yes</td>
<td>CBT reduced negative thoughts, which mediated decrease in depression symptoms, but change in negative thoughts did not precede change in outcome for most participants</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stice et al. (2008)</td>
<td>No</td>
<td>Stice et al. (2010)</td>
<td>ATQ (brief)</td>
<td>No</td>
<td>CBT did not reduce negative thoughts, but reduced negative thoughts predicted decreases in depression symptoms</td>
</tr>
<tr>
<td></td>
<td>TADS Team (2004)</td>
<td>No</td>
<td>Jacobs et al. (2009)</td>
<td>DAS Perfectionism scale</td>
<td>No</td>
<td>CBT did not decrease perfectionism, treatment effect across four treatment conditions decreased when perfectionism change scores were included in model, but effect size of CBT did not appear different with inclusion of these change scores</td>
</tr>
</tbody>
</table>
Table 1.3 (Continued).

<table>
<thead>
<tr>
<th>Outcome Paper</th>
<th>Treatment Conditions</th>
<th>Treatment Effect?</th>
<th>Mediation Paper</th>
<th>Candidate Mediator</th>
<th>Significant Mediator?</th>
<th>Finding (treatment effects described reflect comparison to control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yu &amp; Seligman (2002)</td>
<td>CBT group v. Waitlist</td>
<td>Yes</td>
<td>Yu &amp; Seligman (2002)</td>
<td>Children's Attributional Style Questionnaire (CASQ) - Negative Style scale</td>
<td>Yes</td>
<td>CBT reduced negative explanatory style, which mediated decrease in depression symptoms</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CASQ - Positive Style scale</td>
<td>No</td>
<td>CBT did not increase positive explanatory style</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Difference between CASQ Positive and Negative Explanatory Style scores</td>
<td>Yes</td>
<td>CBT reduced negative explanatory style relative to positive explanatory style, which mediated decrease in depression symptoms</td>
</tr>
</tbody>
</table>

**Social Skills and Relations**

<table>
<thead>
<tr>
<th>Outcome Paper</th>
<th>Treatment Conditions</th>
<th>Treatment Effect?</th>
<th>Mediation Paper</th>
<th>Candidate Mediator</th>
<th>Significant Mediator?</th>
<th>Finding (treatment effects described reflect comparison to control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rohde et al. (2004)</td>
<td>CBT group v. Case management</td>
<td>Yes</td>
<td>Kaufman et al. (2005)</td>
<td>Pleasant Activities Schedule (PES) - Social Skills scale</td>
<td>No</td>
<td>CBT did not improve social skills</td>
</tr>
<tr>
<td>Stice et al. (2008)</td>
<td>CBT group v. No treatment</td>
<td>Yes</td>
<td>Stice et al. (2010)</td>
<td>Loneliness Scale (brief)</td>
<td>No</td>
<td>CBT reduced loneliness but authors did not continue mediation test</td>
</tr>
</tbody>
</table>

**Family Dysfunction**

<table>
<thead>
<tr>
<th>Outcome Paper</th>
<th>Treatment Conditions</th>
<th>Treatment Effect?</th>
<th>Mediation Paper</th>
<th>Candidate Mediator</th>
<th>Significant Mediator?</th>
<th>Finding (treatment effects described reflect comparison to control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brent et al. (1997)</td>
<td>CBT individual v. Attention placebo</td>
<td>Yes</td>
<td>Dietz et al. (2014)</td>
<td>Family Interaction Coding System (FICS) - Dyadic Conflict scale</td>
<td>No</td>
<td>CBT did not reduce observed family conflict</td>
</tr>
<tr>
<td>Kolko et al. (2000)</td>
<td></td>
<td></td>
<td></td>
<td>Family Assessment Device (FAD) - Affective Involvement, Affective Responsiveness, Behavioral Control, Communication, General Functioning, and Roles scales (parent-report)</td>
<td>No</td>
<td>CBT reduced parent-reported family dysfunction on FAD General Functioning and Behavioral Control scales, but not on any other scale; authors did not continue to test for mediation as treatment effect was nonsignificant in mediation paper</td>
</tr>
<tr>
<td>Outcome Paper</td>
<td>Treatment Conditions</td>
<td>Treatment Effect?</td>
<td>Mediation Paper</td>
<td>Candidate Mediator</td>
<td>Significant Mediator?</td>
<td>Finding (treatment effects described reflect comparison to control)</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>----------------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Rohde et al. (2004)</td>
<td>CBT group v. Case management</td>
<td>Yes</td>
<td>Kaufman et al. (2005)</td>
<td>FAD - Affective Involvement, Affective Responsiveness, Behavioral Control, Communication, General Functioning, and Roles scales (youth-report)</td>
<td>No</td>
<td>CBT did not reduce youth-reported family dysfunction on any FAD scale</td>
</tr>
<tr>
<td>Areas of Change Questionnaire (ACQ; parent-report)</td>
<td>No</td>
<td>CBT did not reduce parent-reported family relationship problems</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACQ (youth-report)</td>
<td>No</td>
<td>CBT did not reduce youth-reported family relationship problems</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conflict Behavior Questionnaire (CBQ; parent-report)</td>
<td>No</td>
<td>CBT did not reduce parent-reported family conflict</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBQ (youth-report)</td>
<td>No</td>
<td>CBT did not reduce youth-reported family conflict</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT did not improve family dysfunction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Problem Solving</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brent et al. (1997)</td>
<td>CBT individual v. Attention placebo</td>
<td>Yes</td>
<td>Kolko et al. (2000)</td>
<td>FAD - Problem Solving (parent-report)</td>
<td>No</td>
<td>CBT did not reduce parent-reported family problem solving</td>
</tr>
<tr>
<td>FAD - Problem Solving (youth-report)</td>
<td>No</td>
<td>CBT did not reduce youth-reported family problem solving</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dietz et al. (2014)</td>
<td>FICS - Adolescent Problem Solving scale</td>
<td>Yes</td>
<td>CBT improved observed adolescent problem solving, which predicted higher rates of remission; test of indirect effect was significant only for youth with mothers whose depression symptoms were low to moderate at pretreatment (moderated mediation)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.3 (Continued).

<table>
<thead>
<tr>
<th>Outcome Paper</th>
<th>Treatment Conditions</th>
<th>Treatment Effect?</th>
<th>Mediation Paper</th>
<th>Candidate Mediator</th>
<th>Significant Mediator?</th>
<th>Finding (treatment effects described reflect comparison to control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TADS Team (2004)</td>
<td>CBT individual v. Pill placebo</td>
<td>No</td>
<td>Lewis et al. (2009)</td>
<td>Stage of Change Questionnaire (SOCQ) – Action scale</td>
<td>Yes</td>
<td>CBT increased action stage of change, which mediated decreases in depression symptoms across four treatment conditions, but unclear if mediation specific to CBT only vs. control comparison is present</td>
</tr>
<tr>
<td><strong>Pleasant Activities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stice et al. (2008)</td>
<td>CBT bibliotherapy v. No treatment</td>
<td>No</td>
<td>Stice et al. (2010)</td>
<td>PES (brief)</td>
<td>No</td>
<td>CBT did not increase pleasant activities, but increased pleasant activities predicted decreases in depression symptoms</td>
</tr>
<tr>
<td>Stice et al. (2008)</td>
<td>CBT group v. No treatment</td>
<td>Yes</td>
<td>Stice et al. (2010)</td>
<td>PES (brief)</td>
<td>Yes</td>
<td>CBT increased pleasant activities, which mediated decrease in depression symptoms, but change in negative thoughts did not precede change in outcome for most participants</td>
</tr>
<tr>
<td>Rohde et al. (2004)</td>
<td>CBT group v. Case management</td>
<td>Yes</td>
<td>Kaufman et al. (2005)</td>
<td>PES - Relaxation scale</td>
<td>No</td>
<td>CBT did not increase relaxation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CBT did not increase pleasant activities</td>
</tr>
<tr>
<td><strong>Avoidant Coping</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TADS Team (2004)</td>
<td>CBT individual v. Pill placebo</td>
<td>No</td>
<td>Lewis et al. (2009)</td>
<td>SOCQ - Precontemplation scale</td>
<td>No</td>
<td>CBT did not change precontemplation</td>
</tr>
<tr>
<td><strong>Role Engagement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brent et al. (1997)</td>
<td>CBT individual v. Attention placebo</td>
<td>Yes</td>
<td>Dietz et al. (2014)</td>
<td>FICS - Adolescent Involvement</td>
<td>No</td>
<td>CBT did not reduce observed adolescent involvement</td>
</tr>
</tbody>
</table>
Table 1.3 (Continued).

<table>
<thead>
<tr>
<th>Outcome Paper</th>
<th>Treatment Conditions</th>
<th>Treatment Effect?</th>
<th>Mediation Paper</th>
<th>Candidate Mediator</th>
<th>Significant Mediator?</th>
<th>Finding (treatment effects described reflect comparison to control)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Motivation to Change</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TADS Team (2004)</td>
<td>CBT individual v. Pill placebo</td>
<td>No</td>
<td>Lewis et al. (2009)</td>
<td>SOCQ - Contemplation scale</td>
<td>No</td>
<td>CBT did not change contemplation</td>
</tr>
<tr>
<td><strong>Maintenance of Gains</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TADS Team (2004)</td>
<td>CBT individual v. Pill placebo</td>
<td>No</td>
<td>Lewis et al. (2009)</td>
<td>SOCQ - Maintenance scale</td>
<td>No</td>
<td>CBT did not change maintenance</td>
</tr>
<tr>
<td><strong>Therapeutic Alliance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rohde et al. (2004)</td>
<td>CBT group v. Case management</td>
<td>Yes</td>
<td>Kaufman et al. (2005)</td>
<td>Working Alliance Inventory (WAI)</td>
<td>No</td>
<td>CBT improved alliance compared to control, but improved alliance did not predict decreases in depression symptoms</td>
</tr>
<tr>
<td><strong>Expression of Affect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stice et al. (2008)</td>
<td>CBT group v. No treatment</td>
<td>Yes</td>
<td>Stice et al. (2010)</td>
<td>Emotional Expression</td>
<td>No</td>
<td>CBT increased emotional expression</td>
</tr>
<tr>
<td><strong>Group Cohesiveness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rohde et al. (2004)</td>
<td>CBT group v. Case management</td>
<td>Yes</td>
<td>Kaufman et al. (2005)</td>
<td>Group Cohesiveness Questionnaire</td>
<td>No</td>
<td>CBT did not improve group cohesiveness</td>
</tr>
</tbody>
</table>
Table 1.3 (Continued).

<table>
<thead>
<tr>
<th>Outcome Paper</th>
<th>Treatment Conditions</th>
<th>Treatment Effect?</th>
<th>Mediation Paper</th>
<th>Candidate Mediator</th>
<th>Significant Mediator?</th>
<th>Finding (treatment effects described reflect comparison to control)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent Marital Satisfaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brent et al. (1997)</td>
<td>CBT individual v. Attention placebo</td>
<td>Yes</td>
<td>Kolko et al. (2000)</td>
<td>Locke-Wallace Marital-Adjustment Test (parent-report)</td>
<td>No</td>
<td>CBT improved marital satisfaction, but authors did not continue mediation test as treatment effect was nonsignificant in mediation paper</td>
</tr>
</tbody>
</table>

Note. CBT = cognitive behavioral therapy. aWe categorized the Issues Checklist as Family Dysfunction rather than Problem Solving as conceptualized by Kaufman et al. (2005) because the items are similar to other measures of family dysfunction. bWe categorized SOCQ Action scale as Problem Solving rather than Motivation to Change because the items were similar to other measures of problem solving.
Next most commonly assessed as mediators are social skills and relations, family dysfunction, problem solving, and pleasant activities—each assessed by two RCTs. Kaufman et al. (2005) and Stice et al. (2010) conducted three mediation tests on measures of social skills and loneliness (although Stice et al., 2010, appear not to have completed testing for the CBT group)—none yielded significant findings. Family dysfunction was also not identified as a significant treatment mediator despite the testing of numerous measures of various aspects of family dysfunction (e.g., family affective involvement, family communication, family conflict behavior, general family functioning) with various informants (Dietz et al., 2014; Kolko et al., 2000; Kaufman et al., 2005). Four measures of problem solving were assessed—two family problem solving measures did not mediate outcome improvement, but the other two did. One mediator, the action stage of change, although conceptualized by study authors (Lewis et al., 2009) as a motivational construct, was measured by items that closely resemble those in measures of problem solving (e.g., “I am actively working on my problems.”) and was therefore categorized in the present review as a problem solving measure. The other, adolescent problem solving, significantly mediated rates of remission for youths whose mothers had mild to moderate depression symptoms at pretreatment, but not for youths whose mothers had severe depression (Dietz et al., 2014). This is the only study that demonstrated moderated mediation (i.e., conditional indirect effects; see Preacher, Rucker, & Hayes, 2007) among those reviewed. Four measures of pleasant activity engagement were assessed, all different versions or subsets of the same instrument, and was found to be a significant mediator for CBT group by Stice et al. (2010) but not by Kaufman et al. (2005). It was not identified as a significant mediator for CBT bibliotherapy by Stice et al. (2010).
Categories that were examined by one RCT each as treatment mediators include avoidant coping, role engagement, motivation to change, therapeutic alliance, expression of affect, group cohesiveness, and parent marital satisfaction. All were found to be nonsignificant as mediators. Finally, cognitive skills, treatment credibility, impulsive coping, therapist strategies of any theoretical orientation, research understanding, and perceived stigma were all not analyzed as treatment mediators although they were measured in at least one RCT.

**Study 1A: Discussion**

If we had to use these findings to determine which candidate processes and mechanisms are most promising, which are least promising, and which need to be assessed, we might come up with the following conclusions. First, negative thinking and problem solving have each been identified in at least two RCTs led by different investigators as significantly mediating treatment effects. Therefore, following Chambless and Hollon’s (1998) criteria for identifying empirically supported therapies, we could consider negative thinking and problem solving as empirically supported mediators, and most promising to examine further as change mechanisms. Second, pleasant activities was identified as a mediator in one RCT, and thus would be a possible mediator and next most promising candidate mechanism. Third, all the categories that were assessed as mediators in only one or no RCT need to be further assessed. Fourth, social skills and relations and family dysfunction were each found not to be significant mediators in two RCTs and are thus the least promising candidate mechanisms of youth depression EBPs.

On the other hand, these conclusions feel overly simplistic, and give rise to a number of questions. How representative are these six RCTs of the evidence base of CBT for youth depression? CBT in individual, group, and bibliotherapy formats are represented, as are a variety of control conditions (waitlist, attention placebo, pill placebo, case management). However, none
of the RCTs tested BT or IPT. In addition, the evidence on all three “promising” candidates is actually mixed, and contradictory evidence should be taken into account, as is now recommended in the identification of EBPs (Southam-Gerow & Prinstein, 2014). But how exactly should contradictory evidence be used to balance supporting evidence? Vote counting (i.e., counting the number of significant vs. nonsignificant effects) is difficult to apply when there are mixed findings within studies, and does not incorporate information on between-study differences in the precision or size of the effect, or in participant, treatment, or design characteristics. Indeed, vote counting is recommended only as a last resort (Deeks, Higgins, & Altman, 2011).

Furthermore, how much stock should one put into nonsignificant findings given that individual RCTs are typically underpowered to detect mediation effects (Kraemer et al., 2002)? Moreover, all the mediation analyses involved some form of causal steps approach, which, albeit tremendously helpful in conceptualizing mediation as a series of statistical criteria to be met, has the lowest power among various mediation tests that are currently available (Fritz & MacKinnon, 2007). This is in part because each step requires conducting a separate significance test, and each test carries some degree of error, thus higher powered tests involve a single significance test of the mediation effect to minimize error (Hayes, 2009). Nearly all the reviewed CMs were found to be nonsignificant by virtue of having failed one of the steps. Kolko et al. (2000), for example, did not complete mediation analyses for cognitive distortions and family dysfunction because of a lack of significant treatment effect in the mediation paper, even though the primary outcome paper (Brent et al., 1997) found a significant treatment effect with a different outcome measure. In addition, Kaufman et al. (2005) did not continue testing pleasant activities as a mediator because the treatment effect on this candidate was just above the significance cutoff ($p = .067$).
One wonders how many of these “failed CMs” would have been identified as mediators with a higher powered mediation test. Dietz et al. (2014) were the only authors to employ bias-corrected bootstrapping, which maximizes power to test the significance of the indirect effect (Fritz & MacKinnon, 2007)—but only after following the causal steps approach to eliminate some candidate mediators—thus effectively reducing the power of mediation analyses for those candidates. Quantitative researchers (e.g., Hayes, 2009) recommend the bias-corrected bootstrapping approach without going through the causal steps approach, and have developed free user-friendly tools accessible to most researchers (e.g., Hayes, 2013 PROCESS macro for SPSS or SAS). Bias-corrected bootstrapping can also be applied with path analysis or SEM to test mediation effects.

We hope that researchers take advantage of the newly developed tool for testing mediation in a more rigorous way. Meanwhile, we continue to face challenges in drawing conclusions from simply reviewing the mediation findings of individual RCTs. Thus we proceed with meta-analysis of a sample of RCTs nearly five times as large as the RCT sample in which authors reported testing for treatment mediators, and with many more measures, to get a clearer, more comprehensive picture on mediation effects among EBPs for youth depression.

**Study 1B: Mediation Meta-Analysis**

The present meta-analysis extends the systematic review findings by quantitatively synthesizing data across RCTs to estimate the overall magnitude, direction, and significance of mediation effects for each selected CM. We used an emerging methodology known as model-based meta-analysis (Becker, 2009) or meta-analytic structural equation modeling (MASEM; Cheung & Hafdahl, 2016), which combines meta-analysis and structural equation modeling (SEM) techniques—empirical findings are first synthesized into a pooled correlation or
covariance matrix, and then the matrix is used to fit the models of interest. We pooled correlations among treatment condition, CM, and outcome at various time points across RCTs; then we tested whether the mediation model shown Figure 1.2 fit the pooled correlation matrix (see Figure 1.2). This model tests whether posttreatment levels of the CM mediated the effects of treatment condition on posttreatment symptom outcomes, controlling for pretreatment levels of the CM and symptoms. This approach is similar to assessing whether differential change in CM due to treatment mediates change in symptoms, but avoids several problems of computing pre-post change scores for analyses. For example, change scores are often unreliable (MacKinnon, 2008), are less likely to be reported in journals than scores at each time point, and would increase the time burden of data requests to authors. In addition, we focused analyses on the pre- and posttreatment time points only, and on single mediator models, because data from other time points (e.g., mid-treatment, follow-up), and data that reflected relationships between multiple mediators, were sparse.

Mediation meta-analysis has a number of advantages. First, it enables moving beyond the treatment–CM and CM–outcome relationships analyzed in prior systematic and meta-analytic reviews to directly answer the question of whether a CM mediated treatment effects. Second, it can help resolve the mixed findings resulting from mediation analysis in individual RCTs by estimating a mean mediation effect per CM, while accounting for sample size differences by weighting larger samples more heavily (as they usually yield more precise estimates than smaller samples). Third, it facilitates comparison of different CMs because the conjoint mediational pathway \((ab)\) is measured on a standardized metric (i.e., product of regression coefficients computed from a matrix of pooled correlations that are scale-independent), thus distinguishing stronger CMs from weaker ones. Fourth, existing RCT data are fully exploited because the
analyses can incorporate data from RCTs that did not test mediation, or that provide partial data (i.e., some of the correlations in the matrix)—which make up the bulk of the evidence base. Finally, MASEM can address the lack of power to detect mediation effects in individual RCTs by pooling data from multiple RCTs to increase the total sample of youths.

*Figure 1.2. Two-wave mediation model. The predictor is the treatment condition (e.g., cognitive behavioral therapy vs. waitlist), the mediator is the candidate process or mechanism (e.g., negative thinking) measured at posttreatment, and the outcome is symptom severity (e.g., depressive symptom measure score) measured at posttreatment, with pretreatment measures of the candidate process or mechanism and symptom severity as covariates (i.e., predictors of the mediator variable and the outcome).*

The primary aims of the present meta-analysis are to identify (a) the most robust mediators with the largest effects in the direction consistent with the theoretical framework of the
EBP, which thus merit the labor-intensive and costly next steps required to establish them as change processes and mechanisms; (b) any robust mediators with large effects in the opposite direction, which could generate intriguing questions for future research; (c) candidates that consistently did not mediate treatment outcome, which arguably might have lower priority for further study; and (d) variables with insufficient data that need to be assessed more frequently.

A secondary aim is to demonstrate the application of MASEM to psychotherapy research. The overwhelming majority of meta-analyses deal solely with predictor–outcome/treatment–outcome relationships (see Lipsey & Wilson, 2001); those involving more than two variables are rare in any area (Lipsey, 1997; Shadish, 1996). In the psychotherapy literature, there appear to be only two published meta-analyses of mediation models (which involve at least three variables)—one pertaining to marital and family therapies (Shadish & Sweeney, 1991), and the other to mindfulness-based therapies for adults (Gu, Strauss, Bond, & Cavanagh, 2015). Therefore, this is one of very few meta-analyses of psychotherapy mediation models, and the first one focusing on youths, to our knowledge.

**Study 1B: Method**

The method used to conduct the meta-analysis is the same as that employed for the systematic review, with the addition of effect size extraction and data-analysis, detailed in this section.

**Effect Size Extraction and Computation**

We selected CMs with minimally sufficient data for effect size extraction and subsequent inclusion in the meta-analysis. Though a total sample size of at least 1,000 participants across 10 studies is ideal (Cheung & Chan, 2005), there is no established standard for the minimum
number of studies required for MASEM. Therefore we followed Gu et al.’s (2015) precedent of selecting only CM categories for which data are available for at least three RCTs.

The effect size metric for MASEM is the zero-order correlation, $r$. To fit the mediation model in Figure 1.2, a matrix containing correlations between every pair of variables in the mediation model, to the extent that they were available, was completed for each RCT. Each correlation was extracted by one coder and then checked by at least one other coder. We also coded the number of participants ($n$) used to compute each correlation. Correlation matrices, unlike covariance matrices, are standardized and permit synthesis of findings for different measures of the same CM or outcome (e.g. Becker, 2009). Mediator–outcome relationships were represented by Pearson’s product-moment correlations as both variables are continuous, whereas treatment–mediator and treatment–outcome relationships were represented by point-biserial correlations as one variable is continuous and the other is dichotomous. Pearson’s correlations were most frequently extracted directly from correlation matrices or deidentified datasets, which were routinely requested from RCT authors because the full set of correlations needed are rarely reported. Point-biserial correlations were computed using the following two-step approach. First, Hedges’ $g$ was obtained using the following formulae$^6$ from Lipsey and Wilson (2001; see also Card, 2012),

$$g = \frac{M_t - M_c}{s_{pooled}}$$

(1.1)

where $M_t$ is the mean of treatment group, $M_c$ is the mean of control group, and $s_{pooled}$ is the pooled estimate of the population standard deviation, given by

$^6$ Some authors (e.g., Borenstein, Hedges, Higgins, & Rothstein, 2009b; Lipsey & Wilson, 2001) refer to the statistic computed using these formulae as the standardized mean difference or Cohen’s $d$. Others (e.g, Card, 2012) differentiate between this statistic and another similar statistic in which the denominator is the pooled sample standard deviation by referring to this statistic as Hedges’ $g$, and the other statistic as Cohen’s $d$. 

64
where \( n_t \) is the sample size of the treatment group, \( n_c \) is the sample size of the control group, \( s_t^2 \) is the standard deviation of the treatment group, and \( s_c^2 \) is the standard deviation of the control group. Second, Hedges’ \( g \) was converted to a point-biserial correlation, \( r \), using the following formulae reported in Borenstein et al. (2009a),

\[
 r = \frac{g}{\sqrt{g^2 + a}}
\]

where \( a \) is a correction for unequal sample sizes of the treatment group and control group, given by

\[
a = \frac{(n_t + n_c)^2}{n_t n_c}
\]

Means, standard deviations, and sample sizes of the treatment and control groups were usually reported in articles, or otherwise, requested from RCT authors along with the correlation matrices or deidentified datasets. We also searched articles, dissertations, and theses carefully searched for correlations or other statistics that can be converted to correlations using formulae detailed in Lipsey and Wilson (2001), as implemented via Wilson’s (n.d.) web-based effect size calculator. Pearson’s and point-biserial correlations involving one variable that was not scaled in the default direction were multiplied by -1 so all correlations within a cell, across individual matrices, have the same meaning. When the only information we have about the relationship between two variables is that it is non-significant in a bivariate analysis, we assigned \( r=0 \) (Smith, 1980).
When there were multiple eligible treatment or control conditions, a correlation matrix was created for every unique eligible treatment–control pair within each RCT, separately for each selected mediator category. However, a few RCTs tested multiple eligible EBPs against a control condition, and correlations among CMs and outcomes were available for only the full sample and not for each treatment–control pair—in these cases, the multiple EBP groups were treated as a single treatment group, and had their group means and standard deviations pooled using the formulae detailed by Higgins & Deeks (2011) before computing the point-biserial correlations between treatment condition and CMs or outcomes, so that all the correlations reported for the full sample can be utilized.

The design of most RCTs leads to several sources of statistical dependency within and across correlation matrices. Dependency refers to the problem that data from the same participants or RCT tend to be more closely related compared to data from different participants and RCTs, which could compromise the accuracy of the standard errors of effect size estimates. Thus we addressed each source of dependency in the following ways. First, multiple measures of the same outcome or mediator category within a treatment–control pair contribute more than one correlation to the same cell in the correlation matrix. We took the mean of multiple correlations so that each treatment–control pair contributes only one correlation to each bivariate relationship in the mediation model, following other recent MASEM studies (Carraro & Gaudreau, 2013; Murayama & Elliot, 2012). Second, multiple treatment–control pairs contribute more than one correlation matrix per RCT to the pooled correlation matrix; moreover, the control condition participants tend to be the same in each of the treatment–control pairs. Sophisticated techniques such as multilevel meta-analysis (Van den Noortgate & Onghena, 2003) and robust variance estimation (Hedges, Tipton, & Johnson, 2010) have been developed to address the above sources.
of dependency in meta-analyses of bivariate relationships but are only beginning to be developed for MASEM (Wilson, Polanin, & Lipsey, 2016). Thus we evaluated the impact of including dependent matrices from the same RCT by conducting sensitivity analyses that randomly selected one matrix per RCT, and comparing the results to the original analyses that included dependent matrices. The final source of dependency in our data arises from the same RCT contributing correlations to different cells in the correlation matrix—this is accounted for by the MASEM approach we selected, described in the next section.

**Data Analysis**

We fitted the mediation model depicted in Figure 1.2 to the individual RCT treatment–control pair correlation matrices created for each selected CM category using two-stage structural equation modeling (TSSEM; Cheung, 2014; Cheung & Chan, 2005; for a comparison of MASEM approaches, see Landis, 2013; Zhang, 2011; Cheung & Hafdahl, 2016). In Stage 1, a pooled correlation matrix is synthesized from the individual correlation matrices using multiple-group SEM. In Stage 2, the pooled correlation matrix elements are weighted by their precision using a weighted least squares (WLS) approach and a structural model is fitted to the weighted matrix. There are several advantages to using TSSEM to conduct MASEM (Cheung & Chan, 2005). Besides handling dependency in correlations across cells in the matrix, TSSEM is also equipped to synthesize correlation matrices with missing data (i.e., matrices created from RCTs that collected or reported data on only some of the relationships in the model). Incomplete correlation matrices were included in the present analyses to fully exploit existing data, with maximum likelihood estimates provided by TSSEM for the missing data based on available data in the filled cells. The accuracy and precision of this estimation depends on the amount and nature of available and missing data. In addition, the participant sample size is likely to vary
across correlations in the pooled matrix as at least some studies would not have examined all relationships in the model; TSSEM uses the total sample size (i.e., the sum of participants across all studies), which eliminates ambiguity and arbitrary decisions about choice of sample size (e.g., arithmetic mean, harmonic mean, median) for analyses. To check whether missing data were leading to biased estimates that would change the findings of the present meta-analysis, sensitivity analyses were conducted by rerunning analyses with only complete matrices and examining whether the main findings differ from those generated by the original analyses with complete and incomplete matrices.

TSSEM was conducted using the metaSEM package (Cheung, 2015), which is based on OpenMx (Boker et al., 2011) and implemented in R version 3.3.3. Random-effects models (Cheung, 2014; see also Hedges & Vevea, 1998) were used in both stages whenever possible due to potential heterogeneity among studies examining different EBPs and control groups, recruiting samples with different demographic or clinical characteristics (e.g., age, comorbidity), or using different measures of the same construct. In addition, random effects models allow the MASEM findings to generalize beyond the specific studies analyzed. For each correlation in the pooled matrix, metaSEM computes an absolute measure of heterogeneity, $\tau^2$, the estimated true between-study variance, and a relative measure of heterogeneity, $I^2$, the proportion of observed variance due to true between-study differences (Borenstein, Higgins, Hedges, & Rothstein, 2017; Cheung & Vijayakumar, 2016). In the present analyses, we focused on the $r^2$ because it is a direct measure of how much the correlations vary between studies. Very small heterogeneity variances ($r^2 < 1^{-8}$) were generated for some correlations, leading to estimation problems for one of the mediation models due to limitations of the statistical package in handling such numbers; in this case, we fixed the variances to zero, thus essentially conducting fixed effects analyses for those
highly similar correlations, while allowing larger heterogeneity variances to vary in a user-defined, modified random effects model (Cheung, 2017). The hypothesized mediation models were fitted to each pooled correlation matrix, with tests of direct effects and indirect effects specified. The parameter estimates obtained are the standardized regression coefficients for the relationships between each pair of variables, adjusted for the influence of other variables in the model; 95% likelihood-based confidence intervals (CIs), which offer advantages over Wald approximation-based CIs (Cheung, 2015), indicate whether the standardized regression coefficients are significantly different from zero. Goodness-of-fit indices indicate whether the mediation models fit the data well in an absolute sense, and also relative to one other.

**Study 1B: Results**

Our efforts to retrieve data from the 40 RCTs through various sources (e.g., dissertations, authors) yielded complete data for 18 (45.0%) of the RCTs and partial data for the remainder.

Seven mediator categories had minimally sufficient data for inclusion in the meta-analysis: negative thinking, social skills and relations, family dysfunction, problem solving, cognitive skills, pleasant activities, and avoidant coping. Role engagement was measured by three RCTs at the needed time points, but data were not available from some or all of the three RCTs for some cells in the matrix needed to test the mediation model. Treatment expectancy was measured by only two RCTs at mid- or posttreatment—key time points for testing mediation. All other CMs were each measured by only one or two RCTs across all time points. Mediation models for each of the seven included categories involved all RCTs that administered at least one measure from that category at any time point, and for which at least one correlation involving the mediator category could be computed. Table 1.2 contains details of which RCTs measured which mediator categories, and we note in the text below which of those RCTs were not included in
analyses because correlations involving that mediator category cannot be computed. We also fitted an outcomes-only model, covarying pretreatment levels of the outcome, to the RCTs that were not included in any of the seven mediation models as a gauge of whether these RCTs were comparable to those included in the mediation models in terms of their treatment effects. Table 1.4 displays the pooled correlation estimates for the seven mediation models and outcomes-only model, as well as the range of correlations and number of individual study matrices that went into computing each pooled correlation, and between-study variance of each pooled correlation. We reported the main findings of the most stringent sensitivity analyses that included only complete, independent matrices.

**Negative Thinking**

All 24 RCTs that measured negative thinking were involved in analyses, generating a total of 28 correlation matrices with 2,657 participants. Four of these RCTs (Lewinsohn et al., 1990; Reynolds & Coats, 1986; Rosselló & Bernal, 1999; Stice et al., 2008) contributed two matrices each as they had two eligible treatment conditions, and data were available separately for each treatment–control pair. Another three RCTs (Gillham et al., 2012; Kahn et al., 1990; Stark et al., 1987) had data from their multiple eligible treatment conditions pooled into a single matrix as some data were lacking for separate treatment–control pairs. Only one (Reed, 1994) out of 28 (3.57%) correlation matrices analyzed contained $r=0$ that was coded based on author report of a nonsignificant bivariate relationship; this occurred in the two cells representing the correlations between treatment condition and pretreatment negative thinking, and between treatment condition and pretreatment outcome.
Table 1.4. Pooled Correlation Estimates for the Relationships Between Treatment Condition, Depression Symptoms, and Each Candidate Mediator (CM) at Pre- and Posttreatment

<table>
<thead>
<tr>
<th>CM: Negative Thinking (total ( m=28 ))</th>
<th>Treatment Condition</th>
<th>CM at Pretreatment</th>
<th>Depression Symptoms at Pretreatment</th>
<th>CM at Posttreatment</th>
<th>Depression Symptoms at Posttreatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Condition</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>CM at Pretreatment</td>
<td>( r=0.028, ) (-0.226–0.359), )</td>
<td>---</td>
<td>( m=27, \tau^2=0.009 )</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Depression Symptoms at Pretreatment</td>
<td>( r=-0.010, ) (-0.240–0.350), )</td>
<td>( m=27, \tau^2&lt;1^* )</td>
<td>( m=20, \tau^2=0.062 )</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>CM at Posttreatment</td>
<td>( r=-0.129, ) (-0.445–0.405), )</td>
<td>( m=27, \tau^2=0.027 )</td>
<td>( m=18, \tau^2=0.002 )</td>
<td>( m=18, \tau^2=0.005 )</td>
<td>---</td>
</tr>
<tr>
<td>Depression Symptoms at Posttreatment</td>
<td>( r=-0.244, ) (-0.650–0.011), )</td>
<td>( r=0.307, ) (-0.127–0.456), )</td>
<td>( m=18, \tau^2=0.003 )</td>
<td>( m=18, \tau^2=0.005 )</td>
<td>( m=18, \tau^2=0.027 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CM: Social Skills and Relations (total ( m=16 ))</th>
<th>Treatment Condition</th>
<th>CM at Pretreatment</th>
<th>Depression Symptoms at Pretreatment</th>
<th>CM at Posttreatment</th>
<th>Depression Symptoms at Posttreatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Condition</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>CM at Pretreatment</td>
<td>( r=-0.027, ) (-0.177–0.150), )</td>
<td>---</td>
<td>( m=16, \tau^2&lt;1^* )</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Depression Symptoms at Pretreatment</td>
<td>( r=-0.032, ) (-0.148–0.173), )</td>
<td>( m=16, \tau^2&lt;1^* )</td>
<td>( m=12, \tau^2=0.011 )</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>
Table 1.4 (Continued).

<table>
<thead>
<tr>
<th>Treatment Condition</th>
<th>CM at Posttreatment</th>
<th>Depression Symptoms at Pretreatment</th>
<th>Depression Symptoms at Posttreatment</th>
<th>CM: Family Dysfunction (total m=13)</th>
<th>CM: Problem-Solving (total m=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r=0.067, r=0.548, r=0.160,</td>
<td>-0.206–0.263, 0.313–0.691, -0.295–0.266,</td>
<td>-0.586–0.003, -0.305–0.198, 0.158–0.669, -0.740–0.174,</td>
<td>r=0.007, (-0.127–0.422),</td>
<td>r=0.028, (-0.148–0.038),</td>
</tr>
<tr>
<td></td>
<td>(-0.206–0.263), m=15, r²=0.003</td>
<td>(-0.206–0.263), m=11, r²=0.013</td>
<td>(-0.206–0.263), m=11, r²=0.013</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>r=0.014, r=0.472, r=0.306,</td>
<td>0.001</td>
<td>0.001</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(-0.206–0.263), m=11, r²=0.024</td>
<td>(-0.206–0.263), m=11, r²=0.024</td>
<td>(-0.206–0.263), m=11, r²=0.024</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>
Table 1.4 (Continued).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Condition</th>
<th>CM at Posttreatment</th>
<th>Depression Symptoms at Posttreatment</th>
<th>CM at Pretreatment</th>
<th>Depression Symptoms at Pretreatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM: Cognitive Skills (total m=6)</td>
<td></td>
<td>r=0.038, m=9, ( r^2=0.017 )</td>
<td>r=0.218, m=6, ( r^2=0.034 )</td>
<td>r=0.003, m=7, ( r^2&lt;1 )</td>
<td>r=0.063, m=6, ( r^2=0.059 )</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td>(0.188–0.424), ( \tau^2=0.017 )</td>
<td>(-0.586–0.003), ( \tau^2=0.026 )</td>
<td>(-0.139–0.075), ( \tau^2&lt;1 )</td>
<td>(-0.091–0.045), ( \tau^2=0.037 )</td>
</tr>
<tr>
<td>Condition</td>
<td></td>
<td>r=0.549, ( \tau^2=0.549 )</td>
<td>r=0.549, ( \tau^2=0.549 )</td>
<td>-</td>
<td>r=0.442, ( \tau^2=0.442 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.003–0.866), ( \tau^2=0.055 )</td>
<td>(-0.541–0.119), ( \tau^2=0.034 )</td>
<td>-</td>
<td>r=0.493, ( \tau^2=0.493 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.372–0.375), ( \tau^2=0.034 )</td>
<td>(0.274–0.598), ( \tau^2=0.059 )</td>
<td>-</td>
<td>(-0.003–0.866), ( \tau^2=0.034 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>CM: Pleasant Activities (total m=7)</td>
<td></td>
<td>r=0.036, m=7, ( r^2=0.005 )</td>
<td>r=0.030, m=5, ( r^2=0.026 )</td>
<td>r=0.030, m=7, ( r^2&lt;1 )</td>
<td>r=0.030, m=7, ( r^2&lt;1 )</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td>(0.273–0.052), ( \tau^2=0.005 )</td>
<td>(-0.091–0.045), ( \tau^2&lt;1 )</td>
<td>(-0.091–0.045), ( \tau^2&lt;1 )</td>
<td>(-0.091–0.045), ( \tau^2&lt;1 )</td>
</tr>
<tr>
<td>Condition</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 1.4 (Continued)

<table>
<thead>
<tr>
<th>Treatment Condition</th>
<th>CM at Posttreatment</th>
<th>Depression Symptoms at Pretreatment</th>
<th>Depression Symptoms at Posttreatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r=0.014, (-0.132–0.208)$, $m=7, \tau^2=0.006$</td>
<td>$r=0.553, (0.121–0.755)$, $m=5, \tau^2=0.021$</td>
<td>$r=-0.146, (-0.215–0.032)$, $m=5, \tau^2&lt;1^{-8}$</td>
</tr>
<tr>
<td></td>
<td>$r=-0.236, (-0.586–0.051)$, $m=7, \tau^2=0.019$</td>
<td>$r=-0.146, (-0.225–0.065)$, $m=5, \tau^2&lt;1^{-8}$</td>
<td>$r=0.548, (0.471–0.669)$, $m=6, \tau^2=0.003$</td>
</tr>
<tr>
<td></td>
<td>$r=-0.114, (-0.228–0.080)$, $m=3, \tau^2=0.009$</td>
<td>$r=0.368, (0.359–0.483)$, $m=3, \tau^2=0.008$</td>
<td>$r=-0.048, (-0.387–0.108)$, $m=3, \tau^2=0.022$</td>
</tr>
<tr>
<td></td>
<td>$r=-0.139, (-0.372–0.003)$, $m=3, \tau^2=0.016$</td>
<td>$r=0.031, (-0.097–0.084)$, $m=3, \tau^2&lt;1^{-8}$</td>
<td>$r=0.453, (0.274–0.519)$, $m=3, \tau^2=0.026$</td>
</tr>
</tbody>
</table>

**CM: Avoidant Coping** (total $m=3$)

<table>
<thead>
<tr>
<th>Treatment Condition</th>
<th>CM at Pretreatment</th>
<th>Depression Symptoms at Pretreatment</th>
<th>---</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r=-0.051, (-0.087–0.083)$, $m=3, \tau^2&lt;1^{-8}$</td>
<td>$r=0.059, (-0.089–0.066)$, $m=3, \tau^2=0.008$</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>$r=-0.114, (-0.228–0.080)$, $m=3, \tau^2=0.009$</td>
<td>$r=0.368, (0.359–0.483)$, $m=3, \tau^2=0.008$</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>$r=-0.139, (-0.372–0.003)$, $m=3, \tau^2=0.016$</td>
<td>$r=0.031, (-0.097–0.084)$, $m=3, \tau^2&lt;1^{-8}$</td>
<td>---</td>
</tr>
</tbody>
</table>

**Outcomes Only** (total $m=19$)

<table>
<thead>
<tr>
<th>Treatment Condition</th>
<th>Depression Symptoms at Pretreatment</th>
<th>---</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r=-0.033, (-0.035–0.143)$, $m=18, \tau^2&lt;1^{-8}$</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>$r=-0.066, (-0.280–0.287)$, $m=19, \tau^2=0.011$</td>
<td>$r=0.461, (0.261–0.552)$, $m=14, \tau^2=0.003$</td>
</tr>
</tbody>
</table>
Table 1.4 (Continued).

*Note:* \( r \) = pooled mean correlation, ( ) = range of correlations pooled, \( m \) = number of study matrices contributing to that cell, \( \tau^2 \) = the between-study variance.
Stage 1 analyses yielded \( r^2 \) ranging from \(<1^{-8}\) to \(.062\) (median = \(.007\)), suggesting that the paths in the mediation model varied in how much they differed between RCTs. Stage 2 analyses indicated excellent fit of the mediation model to the data \( (\chi^2(4) = 7.05, p = .13, \text{CFI} = 1.00, \text{RMSEA} = .017, \text{SRMR} = .027) \). All model paths were significantly different from zero, including the conjoint mediation pathway critical to mediation: EBPs reduced negative thinking compared to control conditions, and reduced negative thinking predicted less severe depression symptoms at posttreatment (see Figure 1.3a). The indirect effect of treatment condition on outcome via change in negative thinking was significant \( (\beta = -.037, 95\% \text{CI} [-.062, -.017]) \), suggesting that negative thinking is a significant mediator. The direct effect of treatment condition on outcome that does not go through negative thinking was also significant \( (\beta = -.211, 95\% \text{CI} [-.287, -.134]) \), pointing to other untested mediators that account for substantial variance in the treatment–outcome relationship. The indirect effect of negative thinking accounted for 15.0% of the total effect (sum of indirect and direct effects). Sensitivity analyses that included 15 complete, independent matrices showed that the indirect effect remained significant and in the same direction as in the original analyses with incomplete and dependent matrices \( (\beta = -.038, 95\% \text{CI} [-.065, -.019]) \).

**Social Skills and Relations**

Of the 14 RCTs that measured social skills and relations, correlations involving this mediator category were unavailable for 2 RCTs (Asarnow et al., 2002; 2005). The remaining 12 RCTs were included in analyses, producing a total of 16 correlation matrices with 1,223 participants. Four of these RCTs contributed two matrices each as they had two eligible treatment conditions (Lewinsohn et al., 1990; Rosselló & Bernal, 1999; Stice et al., 2008) or two eligible control conditions (Liddle & Spence, 1990), and data were available separately for each
treatment–control pair. One RCT (Stark et al., 1987) had data from two eligible treatments conditions pooled into a single matrix as some data were lacking for separate treatment–control pairs. Only one (Reed, 1994) out of 16 (6.25%) correlation matrices analyzed contained r=0 that was coded based on author report of a nonsignificant bivariate relationship; this occurred in the two cells representing the correlations between treatment condition and pretreatment social skills and relations, and between treatment condition and pretreatment outcome.

Stage 1 analyses yielded \( r^2 \) ranging from \(<1^{-8}\) to .024 (median = .007), suggesting that the paths in the mediation model varied in how much they differed between RCTs. Stage 2 analyses indicated good fit of the mediation model to the data (\( \chi^2(4) = 8.31, p = .08, CFI = .99, RMSEA = .030, SRMR = .034 \)). All model paths were significantly different from zero, including the conjoint mediational pathway: EBPs improved social skills and relations relative to control conditions, and improved social skills and relations predicted less severe depression symptoms at posttreatment (see Figure 1.3b). The indirect effect of treatment condition on outcome via change in social skills and relations was significant (\( \beta = -.012, 95\% CI [-.026, -.003] \)), suggesting that social skills and relations is a significant mediator. The direct effect of treatment condition on outcome was also significant (\( \beta = -.205, 95\% CI [-.282, -.282] \)), pointing to other untested mediators that account for substantial variance in the treatment–outcome relationship. The indirect effect of social skills and relations accounted for 5.6% of the total effect. Sensitivity analyses that included eight complete, independent matrices showed that the indirect effect remained significant and in the same direction as in the original analyses with incomplete and dependent matrices (\( \beta = -.018, 95\% CI [-.040, -.003] \)).
**Family Dysfunction**

Of the 13 RCTs that measured family dysfunction, correlations involving this mediator category were unavailable for 2 RCTs (Asarnow et al., 2005; Gillham et al., 2012). The remaining 11 RCTs encompassed 13 correlation matrices with 1,342 participants. Two of these RCTs (Rosselló & Bernal, 1999; Stice et al., 2008) contributed two matrices each as they had two eligible treatment conditions with data available separately for each treatment–control pair. None of the 13 correlation matrices analyzed contained $r=0$ that was coded based on author report of a nonsignificant bivariate relationship. Stage 1 analyses yielded $r^2$ ranging from $<1^{-8}$ to .028 (median = .009), suggesting that the paths in the mediation model varied in how much they differed between RCTs. Stage 2 analyses indicated excellent model fit to the data ($\chi^2(4) = 0.88, p = .93, CFI = 1.00, RMSEA = 0, SRMR = .009$). The path estimates critical to mediation were not significantly different from 0: EBPs did not reduce family dysfunction relative to control conditions, and family dysfunction was not associated with depression symptoms at posttreatment (see Figure 1.3c). However, EBPs did reduce depression symptoms at posttreatment, controlling for the effects of family dysfunction. The indirect effect of treatment condition on outcome via change in family dysfunction was not significant ($\beta = -.001, 95\% CI [-.007, .003]$), but the direct effect was ($\beta = -.175, 95\% CI [-.261, -.089]$), suggesting that family dysfunction is not a significant mediator, and that there are other untested mediators that account for the treatment–outcome relationship. The indirect effect of family dysfunction accounted for only 0.4% of the total effect. Sensitivity analyses that included nine complete, independent matrices showed that the indirect effect remained nonsignificant and in the same direction as in the original analyses with incomplete and dependent matrices ($\beta = -.003, 95\% CI [-.009, .001]$).
**Problem Solving**

All but one (Asarnow et al., 2002) of the eight RCTs that measured problem solving had available correlations involving this mediator category and were thus included in analyses. The seven RCTs involved a total of nine correlation matrices with 930 participants. Two of these RCTs contributed two matrices each; one had two eligible treatment conditions (Lewinsohn et al., 1990), the other had two eligible control conditions (Liddle & Spence, 1990), and data were available separately for each treatment–control pair. One RCT (Gillham et al., 2012) had data from two eligible treatments conditions pooled into a single matrix as some data were lacking for separate treatment–control pairs. None of the eight correlation matrices analyzed contained \( r = 0 \) that was coded based on author report of a nonsignificant bivariate relationship.

Stage 1 analyses yielded \( r^2 \) ranging from \(<1^{-8}\) to .055 (median = .023), suggesting that the paths in the mediation model varied in how much they differed between RCTs. At stage 2, most indices showed acceptable fit of the mediation model to the data (CFI = .97, RMSEA = .049, SRMR = .040), but one showed poor fit \((\chi^2(4) = 12.83, p = .01)\). The path estimates critical to mediation were not significantly different from 0: EBPs did not improve problem solving relative to control conditions, and problem solving was not associated with depression symptoms at posttreatment (see Figure 1.3d). However, EBPs did reduce depression symptoms at posttreatment, controlling for the effects of problem solving, and greater problem solving ability was associated with less severe depression symptoms at pretreatment. The indirect effect of treatment condition on outcome via change in problem solving was not significant \((\beta = -.006, 95\% \text{ CI} [-.028, .010])\), but the direct effect was \((\beta = -.191, 95\% \text{ CI} [-.320, -.062])\), suggesting that problem solving is not a significant mediator, and that there are other untested mediators that account for the treatment–outcome relationship. The indirect effect of problem solving accounted
for 3.0% of the total effect. Sensitivity analyses that included five complete, independent matrices showed that the indirect effect remained nonsignificant and in the same direction as in the original analyses with incomplete and dependent matrices ($\beta = -.004$, 95% CI [-.029, .007]).

**Cognitive Skills**

All five RCTs that measured pleasant activities were involved in analyses, generating a total of six correlation matrices with 521 participants. One RCT (Lewinsohn et al., 1990) had two eligible treatment conditions with data available separately for each treatment–control pair and thus contributed two matrices. None of the five correlation matrices analyzed contained $r=0$ that was coded based on author report of a nonsignificant bivariate relationship.

Stage 1 analyses yielded $r^2$ ranging from $<1^{-8}$ to .059 (median = .021), suggesting that the paths in the mediation model varied in how much they differed between RCTs. Stage 2 analyses indicated excellent model fit to the data ($\chi^2(4) = 3.85, p = .43, \text{CFI} = 1.00, \text{RMSEA} = 0, \text{SRMR} = .034$). The path estimates critical to mediation were not significantly different from 0: EBPs did not improve cognitive skills relative to control conditions, and cognitive skills was not associated with depression symptoms at posttreatment (see Figure 1.3e). However, EBPs did reduce depression symptoms at posttreatment, controlling for the effects of cognitive skills. The indirect effect of treatment condition on outcome via change in cognitive skills was not significant ($\beta = -.003$, 95% CI [-.026, .014]), but the direct effect was ($\beta = -.190$, 95% CI [-.351, -.029]), suggesting that cognitive skills is not a significant mediator, and that there are other untested mediators that account for the treatment–outcome relationship. The indirect effect of cognitive skills accounted for 1.5% of the total effect. Sensitivity analyses that included four complete, independent matrices, showed that the indirect effect remained nonsignificant and in the same
direction as in the original analyses with incomplete and dependent matrices ($\beta = -.003$, 95% CI [-.033, .021]).

**Pleasant Activities**

All five RCTs that measured pleasant activities were involved in analyses, encompassing a total of seven correlation matrices with 645 participants. Two of these RCTs (Lewinsohn et al., 1990; Stice et al., 2008) had two eligible treatment conditions with data available separately for each treatment–control pair and thus contributed two matrices each. None of the five correlation matrices analyzed contained $r=0$ that was coded based on author report of a nonsignificant bivariate relationship.

Stage 1 analyses yielded $\tau^2$ ranging from <1.8 to .021 (median = .002), suggesting that the paths in the mediation model varied in how much they differed between RCTs. Due to estimation problems in Stage 2 analyses, the five $I^2$ that were near 0 were fixed to 0 in a user-defined modified random effects model and Stage 1 and 2 analyses were repeated. Some indices indicated acceptable model fit (CFI = .97, SRMR = .051) whereas others indicated poor fit ($\chi^2$(4) = 14.27, $p = .006$, RMSEA = .063). One path estimate critical to mediation was not significantly different from 0: EBPs did not increase pleasant activity engagement relative to control conditions (see Figure 1.3f). However, EBPs did reduce depression symptoms at posttreatment, controlling for the effects of pleasant activity engagement, and greater pleasant activity engagement was associated with less severe depression symptoms both at pretreatment and at posttreatment. The indirect effect of treatment condition on outcome via change in pleasant activity engagement was not significant ($\beta = -.010$, 95% CI [-.032, .013]), but the direct effect was ($\beta = -.223$, 95% CI [-.346, -.101]), suggesting that pleasant activity engagement is not significant mediator, and that there are other untested mediators that account for the treatment–
outcome relationship. The indirect effect of pleasant activities accounted for 4.2% of the total effect. Sensitivity analyses that included three complete, independent matrices showed that the indirect effect remained nonsignificant and in the same direction ($\beta = -.010$, 95% CI [-.041, .023]) as in the original analyses with incomplete and dependent matrices; pleasant activity engagement also remained a predictor of posttreatment depression symptoms ($\beta = -.197$, 95% CI [-.286, -.100]).

**Avoidant Coping**

Three of the four RCTs that measured avoidant coping were included in analyses as correlations involving this mediator category were not available for one RCT (Asarnow et al., 2002). The three RCTs produced a total of three correlation matrices with 328 participants. None of the three correlation matrices analyzed contained $r=0$ that was coded based on author report of a nonsignificant bivariate relationship.

Stage 1 analyses yielded $r^2$ ranging from $<1^{-8}$ to .026 (median = .004), suggesting that the paths in the mediation model varied in how much they differed between RCTs. Stage 2 analyses indicated good model fit ($\chi^2(4) = 6.57, p = .16$, CFI = .99, RMSEA = .44, SRMR = .045). The path estimates critical to mediation were not significantly different from 0: EBPs did not reduce avoidant coping compared to control conditions, and avoidant coping was not associated with depression symptoms at posttreatment (see Figure 1.3g). Moreover, EBPs did not reduce depression symptoms at posttreatment, controlling for the effects of avoidant coping. Both the indirect effect ($\beta = -.005$, 95% CI [-.040, .018]) and the direct effect ($\beta = -.087$, 95% CI [-.275, .100]) were not significant. The indirect effect of avoidant coping accounted for 5.4% of the total effect. Sensitivity analyses were not needed as all matrices were complete and independent.
Outcomes Only

Of the 40 RCTs in this systematic review, 12 were not included in any of the seven mediation models because they did not measure a CM at any timepoint (see Table 1.2), they measured CMs that did not fall into the seven selected mediator categories (Wood et al., 1996), or they measured CMs in one or more of the seven categories but correlations related to the CM could not be computed (Asarnow et al., 2005). The 12 RCTs encompassed a total of 19 correlation matrices with 3,414 participants. Two of these RCTs contributed multiple matrices each; one had two eligible treatment conditions (Clarke et al., 1999) and the other had two eligible treatment conditions and three eligible control conditions (Stice et al., 2007), with data available separately for each treatment–control pair. None of the 19 correlation matrices analyzed contained $r=0$ that was coded based on author report of a nonsignificant bivariate relationship.

Stage 1 analyses yielded $\tau^2$ ranging from $<1^{-8}$ to .016 (median = .011), suggesting that the paths in the mediation model varied in how much they differed between RCTs. Stage 2 analyses indicated excellent model fit to the data ($\chi^2(1) = 2.06, p = .16, \text{CFI} = 1.00, \text{RMSEA} = .018, \text{SRMR} = .021$). Consistent with all but one of the mediation models (i.e., avoidant coping), EBPs reduced depression symptoms compared to control conditions, and pretreatment outcomes positively predicted posttreatment outcomes (see Figure 1.3h). However, the treatment effect in this outcomes-only model ($\beta = -0.079$) is less than half that in the six mediation models that demonstrated treatment effects ($\beta = -.176$ to -.248), and very close to that in the only mediation model without a significant treatment effect ($\beta = -.092$). Sensitivity analyses that included nine complete, independent matrices showed that the treatment effect remained significant and in the same direction as in the original analyses with incomplete and dependent matrices ($\beta = -.089$).
Figure 1.3. Paths, indirect effects, and direct effects in the mediation model for each candidate mediator category (a to g), and paths in an outcomes-only model (h) for comparison. Parameter estimates that are significantly different from zero (i.e., 95% likelihood-based confidence intervals do not contain zero) are marked by an asterisk. Paths that are significantly different from zero are depicted as solid lines, whereas nonsignificant paths are depicted as dashed lines.
Figure 1.3 (Continued).

(e) Cognitive skills at pretreatment → Depression symptoms at pretreatment
   Cognitive skills at posttreatment → Depression symptoms at posttreatment
   Treatment condition → Depression symptoms at posttreatment
   Direct effect = -06
   Indirect effect = -0.03
   Total effect = -09
   \[ -08 \]

(f) Pleasant activities at pretreatment → Depression symptoms at pretreatment
   Treatment condition → Depression symptoms at posttreatment
   Direct effect = -04
   Indirect effect = -0.01
   Total effect = -03
   \[ -31\]

(g) Avoidant coping at pretreatment → Depression symptoms at pretreatment
   Avoidant coping at posttreatment → Depression symptoms at posttreatment
   Treatment condition → Depression symptoms at posttreatment
   Direct effect = -08
   Indirect effect = -0.05
   Total effect = -09
   \[ -09 \]

(h) Treatment condition → Depression symptoms at posttreatment
   Direct effect = -09
   Total effect = -08
Study 1B: Discussion

The present meta-analysis yielded some results that are consistent with those from the systematic review, and other results that were contradictory. Consistent with the systematic review, negative thinking emerged as a significant treatment mediator; family dysfunction and avoidant coping did not. Contrary to the systematic review, social skills and relations also emerged as a significant mediator, whereas problem solving did not, and pleasant activity engagement was identified as a predictor but not a mediator. Finally, cognitive skills, which was not subjected to mediation testing in any individual RCT, did not mediate outcomes in the meta-analytic mediation model. Sensitivity analyses indicated that these findings held up even after excluding dependent matrices as well as those with missing data.

In particular, findings for negative thinking, social skills and relations, and family dysfunction may be considered the most reliable as they were based on a pooled sample size of more than 1,000 participants from 24, 12, and 11 RCTs respectively. Negative thinking and social skills and relations may be considered robust mediators of EBPs for youth depression. Notably, negative thinking accounted for nearly three times the proportion of variance of the total effect of treatment condition on outcome compared to social skills and relations (15.0% vs. 5.6%). Although negative thinking appears to play a pretty large role relative to social skills and relations, its role seems modest from an absolute perspective. The pooled point-biserial correlation of treatment condition with posttreatment negative thinking is only -.13, which translates roughly to $d = 0.26$, considered a small effect by Cohen’s (1988) standards, and even smaller than the $d = 0.35$ found by Chu & Harrison (2007) with a smaller set of studies. In addition, family dysfunction consistently failed to mediate the effects of EBPs on youth depression relative to control conditions. We note, however, that EBPs did not significantly
reduce family dysfunction, which lines up with most CBT protocols’ focus on building the youths’ skills, with only brief parent check-ins to facilitate skill-building in the youth. Therefore, we may conclude that family dysfunction is not a significant mediator, and is unlikely to be a change mechanism, of existing CBT, IPT, and BT protocols. However, it is entirely plausible that family dysfunction might be a mediator and change mechanisms of other therapies that have been or could be developed to effectively target this CM. Thus the status of family dysfunction as a CM of psychotherapies for youth depression in general could be considered to be inconclusive. No CM examined was found to be a robust mediator in the opposite direction favoring the control condition.

There are several plausible, non-mutually exclusive explanations for the discrepant findings between the systematic review and the meta-analysis. First, MASEM might have increased the power to detect mediation effects for some categories, as suggested earlier, while still suffering from inadequate power for other categories. Social skills and relations was analyzed as a mediator in only two individual RCTs, but data from 12 RCTs were pooled in the model that demonstrated it to be a significant mediator. Problem solving and pleasant activities, each analyzed in two individual RCTs as well, had data pooled from seven and five RCTs in their respective mediation models. However, the total sample size in both models had less than the ideal sample size of 1,000 participants, despite persistent efforts by our team to obtain data from all eligible RCTs. Thus those models may well have suffered from a lack of power to detect mediation effects. These two models examining problem solving and pleasant activities as mediators also showed poor fit to the data on some indices, suggesting that these two CMs may relate to treatment condition and outcome in different way from what is specified in the mediation model. Second, some of the individual RCT mediation findings could have been
chance findings that were published in an academic system that values significant findings over
nonsignificant findings. By including CMs that were administered but not tested as mediators in
published reports, our meta-analysis likely mitigated publication bias to some extent and also
increased the reliability of estimates. Third, our mediator categories could have been overly
broad and masked differences between specific measures within a category. Thus specific
measures of pleasant activities and problem solving found to be significant mediators by RCT
authors could have had their effects masked by pooling with other measures we coded in the
same category. Indeed, all seven mediation models had consider heterogeneity variance on some
of the paths.

Finally, it is interesting that the RCTs in the avoidant coping and outcomes-only models
had much smaller treatment effects than the RCTs in the other mediation models. The avoidant
coping model only included three studies in the analyses, one of which is the large multisite trial,
TADS, which did not find a significant effect of CBT vs. pill placebo, and likely attenuated both
the total effect and indirect effect in this model. We note that the proportion of total variance
explained by avoidant coping (5.4%) is nearly as much as that explained by social skills and
relations (5.6%), except that the RCTs in the social skills and relations model (β = -.22) had a
much larger total effect than those in the avoidant coping model (β = -.09). As for the outcomes-
only RCTs, it is possible that they did in fact evaluate CMs, found them to be nonsignificant as
outcomes or mediators, which is more likely given that the treatment effect was smaller, and then
left these mediation studies unpublished in the file-drawer. We attempted to minimize this
possibility by asking the authors of all RCTs to provide information and data about possible CMs
that were not reported in published articles, dissertations, or theses; and all but one RCT
analyzed in this model were purported not to have measured any possible CMs (and not merely
lack data for the CM). Another possible explanation could be that RCTs that lacked the resources to measure CMs also tended to be less rigorously designed and carried out, leading to weaker treatment effects.

**Study 1: General Discussion**

To accelerate the identification of change processes and mechanisms for youth depression, we conducted a systematic review and a meta-analysis of mediation effects in 40 RCTs of EBPs for youth depression published over the past 30 years (up through 2013).

Our systematic review revealed that 75% of RCTs measured some candidate process or mechanism, but only 15%—six RCTs, all focused on CBT—analyzed any of them as mediators. Negative thinking and problem solving were each found to be significant mediators in at least two RCTs conducted by independent research teams, and pleasant activities was a significant mediator in one RCT. No other significant mediators were found, including among measures of social skills and relations and family dysfunction that were tested as mediators in two RCTs. All other categories were tested in either one RCT or not at all, and require further study.

Our meta-analysis enabled us to address some inherent weaknesses of the systematic review by synthesizing data across RCTs to estimate mean mediation effects, by increasing the power of mediation analysis via pooling of RCT samples and use of a single significance test of mediation effects, and by nearly quintupling the mediation analysis sample from 15% (6 RCTs) to 70% (28 RCTs) of the youth depression EBP literature, including RCTs examining IPT and BT. Negative thinking, and social skills and relations, emerged as robust mediators, and family dysfunction was not a significant mediator, drawing from analyses including at least 1,000 participants from 10 RCTs, the optimal sample size for MASEM. Four mediator—categories problem solving, cognitive skills, pleasant activities, and avoidant coping—were also analyzed.
and found not to mediate outcomes, though pleasant activity engagement did predict reduced symptoms; however, the sample size was less than the recommended number, thus these findings could be considered preliminary. It was also notable that, apart from the two robust mediators, the EBPs assessed did not significantly change the remaining CMs relative to controls, thus these CMs might be most appropriately labeled as “inconclusive CMs” of psychotherapy for youth depression in general.

Clinical Implications

The theoretical rationale for CBT is that depression is precipitated or perpetuated by maladaptive thoughts and behaviors (e.g., negative thinking, avoidance), thus changing these thoughts and behaviors to more adaptive ones should lead to symptom reduction (e.g., Clarke & DeBar, 2010). Despite the documented efficacy of CBT for youth depression, the evidence supporting its theory is sparse; as reviewed earlier, the few mediation studies conducted have produced mixed findings. The identification of negative thinking as a robust mediator provides empirical support for the part of this theory relating to changing maladaptive thoughts, and reaffirms that cognitive restructuring, a core part of CBT, may indeed be a key active ingredient in ameliorating depression and should be emphasized in therapy. On the other hand, there is preliminary evidence that pleasant activities and problem solving, the core behavioral change targets in CBT, are not significant mediators of existing EBPs; notwithstanding the fact that sample sizes for these two mediation models are less than ideal, empirical support for the part of CBT relating to changing behaviors remains lacking. One way to interpret this evidence is that there is stronger evidence supporting cognitive change than behavioral change as active ingredients in CBT, and thus argue for prioritizing cognitive restructuring over behavioral activation. Youth psychotherapy researchers have traditionally taken the “toolbox” approach to
treatment development, including a variety of skills and components in a protocol that the therapist and youth test out before picking the ones that work best. More recent efforts have instead focused on slimming down these bulky toolboxes to improve the disseminability of evidence-based practices in everyday clinical settings (see Weisz, Bearman, Santucci, & Jensen-Doss, 2017).

However, the true picture of how CBT improves outcomes is likely more nuanced than that described previously. Support for the cognitive change part of CBT theory would have been stronger had cognitive skills emerged as a mediator. First, measuring cognitive skills is a more direct way to assess the cognitive reframing skills that CBT purports to foster, than measuring negative thinking. Second, compared with negative thinking, cognitive skills has less conceptual overlap with depression symptoms. With only five RCTs contributing data to the cognitive skills model, the analyses could have been underpowered, or the types of cognitive skills measured may not have been particularly potent, or the measures are not reliably tapping into the cognitive skills that youths develop through CBT. In any case, more work would be needed to elucidate which cognitive skills ameliorate negative thinking and depression symptoms, and how they achieve this.

More research is also needed to clarify the role of pleasant activity engagement in improving outcomes. The present meta-analysis showed that pleasant activities predicted better outcome, but EBPs did not increase pleasant activities relative to control conditions. Also, the proportion of the total effect accounted for by the indirect effect through pleasant activities (4.2%) to close to that accounted for by one of the robust mediators, social skills and relations (5.6%). Does this suggest that pleasant activity engagement is merely a correlate of depression symptoms, with CBT acting on some other mechanism to change both? Or are existing CBT
protocols not emphasizing pleasant activities scheduling sufficiently, and thus missing out on opportunities to maximize therapeutic benefit? Some researchers are leaning towards the latter scenario, given their ongoing efforts to develop and evaluate behavioral activation (BA) therapy for depressed youth, in which pleasant activity scheduling and problem solving are core components (Chu, Colognori, Weissman, & Bannon, 2009; McCauley et al., 2016). Interestingly, one research team (McCauley et al., 2016) found no significant change in scheduled activity engagement and goal-directed activation in both BA and the comparison condition comprising uncontrolled CBT or IPT, despite symptom reduction in both groups—which supports the present meta-analysis findings that pleasant activities may not be a treatment mediator.

As the other robust mediator, social skills and relations also warrants greater attention. This mediator is a core target of IPT, which posits that positive social relationships are important for emotional wellbeing, thus IPT focuses helping youths building interpersonal skills to improve relationships with peers, romantic partners, or parents in order to reduce their depression symptoms (Jacobson & Mufson, 2010; Rosselló & Bernal, 1999). As we are unaware of any mediation studies conducted with RCTs of IPT for youths, the present meta-analysis may be one of the first mediation studies to provide empirical justification for the theoretical rationale of IPT. Many CBT and BT protocols do not include social skills (for an exception, see Clarke & DeBar, 2010), and in any case, social skills are unlikely to be prioritized over other “tools” including cognitive restructuring, pleasant activity scheduling, and problem solving. The present findings suggest that an increased focus on social skills and relations may boost the therapeutic benefit of CBT and BT. We note, however, that negative thinking explained three times the proportion of the total effect that was explained by social skills and relations in the present meta-analysis—which could mean a few things. First, it is plausible that negative thinking has greater
impact on depression symptoms than social skills and relations. Second, this finding may be an artifact of the RCT sample testing predominantly CBT protocols that do not focus on improving social skills and relationships, thus mediation effects involving social skills and relations are diluted. Third, the social skills and relations mediator category encompasses somewhat heterogeneous measures, including social activity engagement, positive self-presentation, social adjustment, interpersonal relationship quality, social support, loneliness, sociometric ratings of liking, and prosocial behavior. Some of these measures may be mediators, whereas others may not be, and grouping these measures together may lead to a weak overall mediation effect.

Limitations

Chief among the present review’s limitations is the issue of temporal precedence of the CM—the Achilles’ heel of research on mediators, mechanisms, and processes. Indeed, Stice et al.’s (2010) mediation findings suggest that for most participants, change in negative thoughts do not occur before symptom reduction. Our mediation models were constrained by the data available at each time point; restricting the CM data to the midtreatment time point would have severely restricted our sample size. A less severe sample size restriction would result from including posttreatment CM data and follow-up outcome data in the model, but the results from this model would have a substantively different interpretation from one where the outcome is measured at posttreatment. In either case, neither of the above models would preclude the outcome from changing concurrently with, or before the CM. More intensive measurements at multiple time points during treatment would be required to build the kind of model that can establish temporal precedence of the CM, but this measurement schedule is rarely utilized in youth psychotherapy RCTs (for an exception, see Marker, Comer, Abramova, & Kendall, 2013). Thus we offer the caveat that negative thinking and social skills and relations are robust
mediators according to our findings, but their temporal precedence needs to be determined in order to meet criteria for change mechanism status.

Other limitations are sample-related. Although we included non-English-language sources, dissertations, theses, and unpublished data related to our sample of 40 RCTs, the RCTs themselves were drawn from our existing meta-analytic database that includes only outcome studies published in peer-reviewed English language journals. Thus the RCTs are vulnerable to publication bias, as the published outcome papers are more likely to have been selected for significant findings than non-English journal articles, dissertations, or theses. RCTs published from 2014 were also omitted from this database, a restrictions that is compounded by the small number of RCTs included compared to the majority of published meta-analyses of bivariate relationships. Thus, adding a few more recent RCTs could very well change the results of some of the mediation models; and the relatively few RCTs in each mediation model may not generalize to the full range of RCTs of EBPs for youth depression, much less to EBPs as they are carried out in usual clinical care settings. Indeed, with a few exceptions (Weisz et al., 2009), the vast majority of included RCTs appeared to be efficacy trials conducted in research settings, with study therapists and recruited youths, rather than the more clinically representative effectiveness trials conducted in practice settings, with practicing clinicians and referred youths—which unfortunately is not a unique problem to youth depression treatments (see Weisz, Ng, & Bearman, 2014).

Further limitations relate to statistical dependency. Our handling of dependence among multiple treatment or control groups from the same RCT (i.e., including each unique treatment–control pair), and among multiple outcomes or CMs of the same category, from the same RCT (i.e., taking the mean), although commonly undertaken, was not ideal. Although sensitivity
analyses revealed that excluding dependent matrices did not change the overall conclusions of the present meta-analysis, sensitivity analyses were not conducted to exclude dependent correlations within the same cell due to multiple outcomes or CMs. More sophisticated approaches that incorporate multilevel meta-analysis into TSSEM to account for dependent correlations (Wilson et al., 2016) may be helpful for addressing dependency in the present analyses.

It bears mentioning that we included eligible RCTs in analyses regardless of whether treatment main effects were found. It is questionable whether those RCTs with negligible treatment effects should in fact be included in analyses. On the one hand, one might argue that EBPs that did not outperform controls, especially passive control conditions such as waitlist, had no treatment effect to be mediated (see Holmbeck, 2009), thus including such RCTs would deflate the overall mediation effect. On the other hand, some of the control conditions contained therapeutic elements (e.g., usual care, case management), and an EBP that did not outperform such a control may have produced improvement in outcome, just not much more than that produced by the control. Knowing the candidate processes and mechanisms by which the EBP improved outcomes would still be valuable. Moreover, some RCTs may have small treatment effects that did not attain significance, but could contribute to an overall main treatment effect once data from multiple RCTs are pooled. Sensitivity analyses may provide useful information about whether mediation effects are strengthened once RCTs with passive controls and no treatment effects are removed from the model, if a sufficient number of RCTs remain to be analyzed.

Some of the above limitations reflect the reality of the field, such as the small number of studies, the low frequency at which CMs and outcomes were measured during treatment, and the
difficulty of obtaining data for all the bivariate relationships in the model for all studies. As reported earlier, nearly all authors responded but several simply did not have the time or resources to retrieve the data needed, despite our attempts to contribute resources to assist in data retrieval. Sometimes the data were irretrievable, having been stored on very old equipment predating the ubiquitous use of personal computers. Other limitations (e.g., excluding RCTs with primary outcome papers that are not published in English-language peer-reviewed journals) could possibly be addressed with additional work. We consider this attempt as a working model for one way to move forward from the scant and mixed findings that individual mediation studies in youth psychotherapy so often produce (see Weisz et al., 2013). We consider the present systematic review and meta-analysis as a working model for one way to move forward from the scant and mixed findings that individual mediation studies in youth psychotherapy have produced to date (see Weisz et al., 2013). We welcome other researchers to improve on our methods and add to the findings, and we will start by suggesting future research directions.

**Future Directions**

Examining moderators can provide information about whether mediation effects differ by EBP type, control type, specific CM measure, specific subset of depression symptoms, and participant characteristics such as age or symptom severity. Research with depressed adults indicates that CBT, IPT and SSRI produce equal benefit when averaged across individuals, but that some particular individuals may do much better with one of these treatments than with the others (DeRubeis et al., 2014; Huibers et al., 2015). In addition, the diagnostic category of depression comprises a heterogeneous cluster of symptoms, such that there is considerable individual variation in the specific constellation of symptoms that may signal different underlying psychopathology, leading to differential treatment response across individuals.
In all probability, the change processes and mechanisms also differ by treatment, participant characteristics, and specific symptoms assessed as the outcome. Unfortunately, testing for moderators in MASEM is much more laborious than testing for moderators in a meta-analysis of bivariate relationships. It is also limited to categorical moderators, which allow clusters or subgroups to be compared. In addition, moderation analysis will further reduce the power of analyses that are already underpowered with a small sample of RCTs, and even larger samples may have few RCTs in any one level of the moderator. Indeed, it can be challenging to balance the need to distinguish between studies and measures with the need to consolidate them to understand overall effects. Nevertheless, it would be worthwhile to at least explore the feasibility of moderator analyses.

Next steps could also involve testing multiple-mediator models, as the direct effect was glaringly larger than the indirect effect in all seven single-mediator models. The relatively small indirect effect, even when significant, indicated that variance of any one mediator accounted only for a small proportion of the variance in outcome. Examining whether multiple mediators jointly account for a more substantial proportion of variance would indicate whether other important mediators remain to be examined. However, the feasibility of testing a multiple mediator model would depend on the number of RCTs with data available for multiple mediator categories, including correlations across categories, and the capability of the MASEM method for synthesizing correlation matrices that have substantial missing data, as RCTs would differ in the categories for which data are available.

Of course, the large direct effect could possibly be accounted for by mediators that were too few in number to be included in analyses, or that were not measured at all in RCTs. All the change processes and common factors did not have sufficient data for analysis, and no change
process or common factor was tested as a mediator in more than one RCT. Also absent from all RCTs were performance-based measures and biological measures; nearly all measures were based on subjective report of the youth, parent, or school staff, or evaluation by research staff, and observation-based measures were rare. Clinical psychological science has developed performance-based measures, particularly cognitive task measures, to elucidate psychopathological processes, and these are beginning to be used in treatment contexts (Cha & DiVasto, 2017). Similarly, candidate biological change mechanisms are increasingly assessed in psychotherapy studies (e.g., Messina, Sambin, Palmieri, & Viviani, 2013). However, these developments appear to have occurred primarily in the adult psychotherapy literature. These gaps in youth depression psychotherapy research could serve as directions for future research. The findings of the present meta-analysis suggest that performance-based and biological measures related to negative thinking, and social skills and functioning, may be promising candidate mechanisms to target.

Conclusion

Reviewing 30 years of RCT research on EBPs for youth depression, we discovered that candidate change processes and mechanisms are measured frequently among RCTs, but seldom tested as treatment mediators. Synthesizing this body of work produced evidence that reduction in negative thinking and improvement in social skills and relations reliably mediated the effects of EBPs for youth depression across studies, and that decreases in family dysfunction reliably failed to do so. These findings help us take stock of what is currently known about how EBPs for youth depression work, and suggest a research agenda for the years ahead.
References

References marked with an asterisk indicate articles, dissertations, and theses included in the present review.


doi:10.1001/jama.293.3.311


http://www.effectivechildtherapy.com


105


doi:10.1037/a0026929


doi:10.1097/01.chi.0000240840.63737.1d


doi:10.1080/15374416.2010.486318


http://dx.doi.org/10.1016/j.jaac.2009.10.007

doi:10.1037/a0014154

University of Sydney, Sydney, Australia.

102. doi:10.1017/S0141347300018218

(Ed.), *Meta-analysis of drug abuse prevention programs* (pp. 216 – 233). NIDA Research

Publications.

Dickerson, J. (2005). Cost-effectiveness of an intervention to prevent depression in at-
doi:10.1001/archpsyc.62.11.1241

Erlbaum Associates.

comparison of methods to test mediation and other intervening variable effects.


*Marchand, E., Ng, J., Rohde, P., & Stice, E. (2010). Effects of an indicated cognitive-behavioral depression prevention program are similar for Asian American, Latino, and European American adolescents. *Behaviour Research and Therapy, 48*(8), 821–825. doi:10.1016/j.brat.2010.05.005


121


*Treatment for Adolescents with Depression Study (TADS) Team. (2004). *Treatment for Adolescents with Depression Study(TADS)/Substance Use and Other Outcomes Following Treatment for Adolescent Depression (SOFTAD) [Data file]. Retrieved from https://ndar.nih.gov/download.html?Collection=2145


*Treatment for Adolescents with Depression Study (TADS) Team. (2007). Treatment for Adolescents With Depression Study (TADS): Long-term effectiveness and safety outcomes. *Archives of General Psychiatry, 64*(10), 1132–1143. doi:10.1001/archpsyc.64.10.1132


doi:10.1177/0145445508322629


doi:10.1080/15374416.2016.1163708


doi:10.1037/0033-2909.132.1.132


Dissertation Study 2:

Assessing Youth Coping and Therapeutic Alliance as Mediators in an Effectiveness Trial

Comparing Cognitive Behavioral Therapy and Usual Care for Youth Internalizing Disorders

Mei Yi Ng

Harvard University
Abstract

**Objective:** Among psychotherapies for internalizing problems (i.e., anxiety and depression), cognitive behavioral therapy has the most evidence. Yet manualized CBT (MCBT) and other evidence-based psychotherapies only confer modest advantages over usual care (UC). To understand the mechanisms through which MCBT and UC work under everyday clinical conditions, we examined treatment mediators among internalizing youths in two effectiveness trials that found comparable outcomes for MCBT and UC. **Method:** Both youths and community therapists were randomized to treatment condition. Therapy session recordings were coded to identify strategies actually used by therapists. We examined improvement in youth coping and therapeutic alliance as mediators of the effects of treatment condition, cognitive strategies, and behavioral strategies on posttreatment internalizing symptoms. The indirect effect was tested with 95% bias-corrected bootstrapped confidence intervals across 50 multiply imputed datasets. **Results:** The only significant mediator was youth-reported secondary control coping (i.e., adjusting oneself to adapt to existing conditions); UC increased secondary control coping, which was associated with lower symptomatology. Youth-reported secondary control coping did not mediate the effects of therapist cognitive or behavioral strategies on outcome. Increases in parent-reported youth secondary control and primary control coping (i.e., influencing existing conditions to fit one’s goals) predicted outcome across treatment conditions. **Conclusion:** UC appears to ameliorate internalizing symptoms by improving youth-reported secondary control coping, and parent-reported primary and secondary control coping may be common factors predicting symptom reduction across treatment conditions.

**Keywords:** children and adolescents, internalizing problems, mediators, cognitive behavioral therapy, usual care
Study 2: Assessing Youth Coping and Therapeutic Alliance as Mediators in an Effectiveness Trial Comparing Cognitive Behavioral Therapy and Usual Care for Youth Internalizing Disorders

Among psychotherapies for youth internalizing problems (i.e., anxiety and depression), cognitive behavioral therapy (CBT) has the largest evidence base. Recent reviews (Higa-McMillan, Francis, Rith-Najaran, & Chorpita, 2016; Weersing, Jeffreys, Do, Schwartz, & Bolano, 2016) show that CBT, including various modalities (e.g., individual, group) and components (e.g., exposure, modeling) have more empirical support than any other therapy for anxious or depressed youths. Yet CBT and other evidence-based psychotherapies (EBPs) only confer—at best—modest advantages over usual care (UC; Spielmans, Gatlin, & McFall, 2010; Wampold et al., 2011, Weisz, Jensen-Doss, & Hawley, 2006; Weisz, Kuppsen, Eckshtain, Ugueto, Hawley, & Jensen-Doss, 2013). The underlying reasons are poorly understood, partly because direct EBP–UC comparisons often share several confounds. The two treatment conditions often differ not only in theoretical orientation but also in flexibility, and therapist vocation. EBP is usually delivered in a manualized format by research therapists and graduate students who have an allegiance to the EBP approach and whose priorities may be to master the EBP techniques and deliver them with high fidelity, with relatively few time and financial constraints; in contrast, UC is usually delivered in an individualized, unstandardized format by practicing clinicians who usually have less of an allegiance to any one approach, and whose priorities may be to treat as many patients as they can through whatever techniques seem most suitable for each individual patient. Indeed, research shows the effectiveness of EBPs tends to decline when delivered under real-world clinical conditions—a phenomenon termed the “implementation cliff” (Weisz, Ng, & Bearman, 2014). The present study was designed to
expand our understanding of how manualized CBT (MCBT) and UC work under real-world clinical conditions, by testing candidate change mechanisms through which the treatments might produce internalizing symptom reduction. The study focused on two candidate mechanisms suggested by prior evidence: youth coping and therapeutic alliance. Both were examined in the context of a study in which MCBT and UC were tested under everyday clinical conditions, and common confounds were addressed in the analysis.

**Youth Primary Control Coping and Secondary Control Coping.**

CBT therapists often teach youths coping strategies that they can use to ameliorate their depressed and anxious feelings. In fact, the word “coping” appears in the names of several CBT protocols for youth depression and anxiety, such as, the Adolescent Coping with Depression Course (CWD-A; Clarke, Lewinsohn, & Hops, 1990) and Coping Cat (CC; Kendall, Kane, Howard, & Siqueland, 1990). Thus youth coping strategies that are consistent with CBT are likely candidates for CBT change mechanisms.

Empirical research shows that two dimensions of coping derived from confirmatory factor analysis are associated with lower levels of internalizing symptoms among youths (Connor-Smith, Compas, Wadsworth, Thomsen, & Saltzman, 2000; see also Weisz, Francis & Bearman, 2010; Weisz, Southam-Gerow, & McCarty, 2001). Primary control involves action taken on people, events, and the environment in order to influence existing conditions and bring them into alignment with one’s wishes and goals, as originally defined by Rothbaum, Weisz, and Synder (1982). Connor-Smith and colleagues (2000) extended the definition of primary control to include not only attempts to influence the environment (e.g., problem-solving) but also to influence one’s emotions (e.g., emotion regulation, emotional expression). On the other hand, secondary control involves altering one’s wishes, expectations, and interpretations in order to
accommodate to existing conditions and bring oneself into alignment with those conditions (Rothbaum et al., 1982; Weisz, Rothbaum, & Blackburn, 1984a, 1984b). Secondary control coping encompasses cognitive restructuring, positive thinking, acceptance, and distraction (Connor-Smith et al., 2000).

Because the two dimensions of coping so closely resemble CBT skills, it is surprising that few randomized controlled trials (RCTs) testing CBT for anxious or depressed youths have measured coping (Chu & Harrison, 2007), and even fewer have tested coping as a mediator of treatment outcome. In one RCT that compared CBT to a life skills tutoring condition for adolescents with depression and comorbid conduct problems, a measure of parent–youth conflict intended to be a proxy for interpersonal problem-solving did not mediate outcomes due to lack of between-treatment differences (Kaufman, Rohde, Seeley, Clarke, & Stice, 2005). In a second RCT, secondary control coping measured at posttreatment mediated effects of family group CBT on youth-reported depression and anxiety among adolescent offspring of parents with a history of major depression at 12-month follow-up compared to a reading materials control condition (Compas et al., 2010). In a third RCT comparing a program for parentally bereaved youths, which included CBT strategies, to a reading materials control condition, positive coping (similar to a combination of primary and secondary control coping) mediated intervention effects on concurrently measured youth internalizing symptoms (Tein, Sandler, Ayers, & Wolchik, 2006). Other RCTs that tested coping as an outcome rather than as a mediator generally seemed to find that CBT improved coping in relation to passive controls such as waitlist (Asarnow, Scott, & Mintz, 2002; Barrett, Dadds, & Rapee, 1996; Flannery-Schroeder & Kendall, 2000; Kendall, 1994; Kendall et al., 1997; King et al., 1998), but not in relation to active controls such as family
education and support because both CBT and the active control improved youth coping to a similar extent (Beidel, Turner, & Morris, 2000; Heyne et al., 2002; Kendall et al., 2008).

The conceptual overlap between distinct sets of CBT skills, on the one hand, and primary and secondary control coping, on the other, suggests that these dimensions of coping may be “specific factors” that drive symptom reduction in CBT (see DeRubeis, Brotman, & Gibbons, 2005). However, the pattern of RCT findings showing that both CBT and active controls (but not passive controls) improve youth coping suggests the possibility that improving youth coping may be a “common factor” of all bona fide therapies (see Wampold, 2005).

**Therapeutic Alliance**

The therapeutic alliance has been conceptualized by Bordin (1979) as the agreement between the patient and therapist on the targeted outcomes of treatment (goal) and how to achieve those outcomes (task), as well as a patient–therapist relationship characterized by positive regard and trust (bond). In youth psychotherapy, both the youth–therapist alliance and the parent–therapist alliance are pertinent as youths are often referred and brought to treatment by parents (Shirk & Karver, 2003).

Therapeutic alliance has been considered a common factor targeted by psychotherapies to improve symptom outcomes regardless of theoretical orientation (Crits-Christoph, Connolly Gibbons, & Mukherjee, 2013), and the evidence is consistent with this notion. Comprehensive meta-analyses of the alliance–outcome relationship have revealed small to medium effects for both adult and youth treatments (Horvath, Del Re, Flückiger, & Symonds, 2013; McLeod, 2011). However, the majority of studies that assessed alliance–outcome relationships among anxious or depressed youths were limited to a single MCBT condition (e.g., Karver, et al., 2008; Labouliere, Reyes, Shirk, Karver, 2015; Shirk, Gudmundsen, Kaplinski, & McMakin, 2008) or multiple
MCBT conditions (e.g., Chiu, McLeod, Har, & Wood, 2009; Hudson et al., 2014; Liber et al., 2010). This limitation precludes investigation of whether different therapies may have differential effects on alliance and whether alliance–outcome relationships may differ by treatment condition.

Among the few studies that assessed alliance–outcome relationships in a MCBT condition and a non-CBT active control condition, findings were mixed. One RCT found that alliance ratings increased over time for both family-based MCBT or family education and support, and that therapist- and mother-rated alliance, but not youth-rated alliance, predicted subsequent anxiety symptom outcome among youths who received either (Marker, Comer, Abramova, & Kendall, 2013). Although consistent with the common factor hypothesis, these findings do not rule out a role for specific factors because differences by treatment condition were not assessed. A second RCT found that alliance predicted anxiety symptom outcome for MCBT but not for sertraline, combined treatment, or placebo (Cummings et al., 2014); this suggests that the alliance–outcome association may not extend to medication or combined psychotherapy–medication treatments, but does not exclude the possibility of alliance being a common factor of therapies without medication. A third RCT found that MCBT resulted in greater youth–therapist alliance than did life skills tutoring, but alliance was not related to depression symptom outcome (Kaufman et al., 2005), which neither supports alliance as a common or specific factor.

In summary, more research is needed to clarify the role of therapeutic alliance in MCBT specifically, and in psychotherapy more generally, for youth internalizing problems, particularly in the context of real-world clinical practice.
The Present Study

The present study was designed to help clarify the potential roles of youth primary control coping and secondary control coping, as well as therapeutic alliance, in driving symptom reduction. This was accomplished through secondary analysis of data from two parallel RCTs that compared MCBT to UC under everyday clinical conditions, known collectively as the Youth Anxiety and Depression Study (YADS; Southam-Gerow et al., 2010; Weisz et al., 2009). YADS recruited treatment-seeking youths at community mental health clinics and the practicing clinicians at those clinics, randomizing both youths and therapists to treatment condition—thereby ensuring that therapeutic orientation would not be confounded with therapist vocation. The strategies actually used by therapists were also observationally coded in both conditions, from therapy session recordings. MCBT was found to include significantly more extensive use of cognitive and behavioral strategies than UC, and both conditions were found to also use psychodynamic, family systems, and client-centered strategies (McLeod & Weisz, 2010; Weisz et al., 2009; Southam-Gerow et al., 2010). The documentation of therapist strategies allows the use of cognitive and behavioral strategies to be separated from the manualized, standardized format of MCBT.

The two RCTs differed in the primary diagnosis of the youths recruited and the MCBT protocol used, with one RCT including youths with a primary anxiety disorder and a protocol that targeted anxiety (Southam-Gerow et al., 2010), and the other including youths with a primary depressive disorder and a protocol that targeted depression (Weisz et al., 2009). Otherwise, the study design and method were essentially identical, the two RCTs were conducted in synchrony, and they produced highly similar results, with substantial improvement overall in diagnosis and symptoms from pre- to posttreatment and no significant between-group
differences in posttreatment outcomes (75% of youths in the depression RCT and 50% of youths in the anxiety RCT no longer met criteria for their primary diagnosis at posttreatment). Given these similarities, and considering that youth anxiety and depression symptoms load onto a common broad internalizing factor (e.g., Achenbach & Rescorla, 2001), and that therapeutic effects for one youth internalizing disorder generalize to the other (Weisz, Weiss, Han, Granger, & Morton; 1995), data from the two RCTs were pooled, with primary diagnosis added as a covariate in the present study.

**Study Aims**

The primary aim of the present study was to evaluate each candidate mechanism as a possible mediator of treatment outcomes. Significant mediation would support that candidate mechanism as a specific factor that is activated more strongly by one treatment condition than the other to reduce symptoms. On the other hand, nonsignificant mediation, but significant prediction of outcome would support that candidate mechanism as a common factor—if both treatment conditions led to equal change in that candidate mechanism over time. It is also plausible that a candidate mechanism may be neither mediator nor predictor, suggesting that it plays an inconsequential role in improving outcome. Some researchers have argued for mediation analysis to be conducted only in the presence of a significant treatment effect (e.g., Baron & Kenny, 1986; Holmbeck, 1997)—which makes perfect sense when comparing a treatment to a passive control—but seems odd when comparing two active treatments that are both intended to produce benefit and that do result in improved outcomes. In fact, according to the specific factors argument, it is probable that two treatments may improve outcomes via targeting different change mechanisms. Many quantitative researchers now recommend
mediation testing regardless of whether main effects are present (e.g., Kraemer, Wilson, Fairburn, & Agras, 2002; Hayes, 2009; MacKinnon, Fairchild, & Fritz, 2007).

A secondary aim is to investigate the relationship between significant treatment mediators and therapists’ use of cognitive strategies or behavioral strategies. Treatment condition could be considered a rough measure of high- vs. low-dose of CBT strategies. Measures of therapist cognitive strategies and behavioral strategies serve as more precise measures of CBT and permit the effects of cognitive strategies to be differentiated from those of behavioral strategies. Thus a second set of analyses were planned to examine if candidate mechanisms that significantly mediated the effects of treatment condition also mediated the effects of therapist cognitive strategies and behavioral strategies. Because it is plausible that MCBT and UC may differ not only in the quantity of CBT strategies used, but that CBT strategies may function in qualitatively different ways in each condition that is not captured by the therapist strategies measure, a final set of analyses were planned for candidate mechanisms that significantly mediated the effects of therapist strategies to test if mediation effects were moderated by treatment condition, primary diagnosis, or both.

**Method**

**Participants**

Youths and therapists from eight community mental health clinics in Los Angeles County participated in the RCTs. Youths were referred through normal community channels; those with a primary diagnosis (via standardized diagnostic interview) of a depressive disorder were included in the depression RCT and those with a primary diagnosis of an anxiety disorder were included in the anxiety RCT. Both youths \( (n = 105) \) and therapists \( (n = 85) \) were assigned to treatment conditions via block randomization procedures (Friedman, Furberg, & DeMets, 1998).
The present study sample includes 91 youths (depression MCBT n = 29, depression UC n = 23, anxiety MCBT n = 18, anxiety UC n = 21) who attended more than one therapy session, and for whom there was at least one posttreatment or follow-up outcome measure completed. These 91 youths were assigned to 80 therapists (MCBT n = 40, UC n = 40), several (8.8%) of whom treated youths in both the depression and anxiety RCTs within their assigned treatment condition. The majority (76.3%) of therapists treated one youth in the present sample; the other therapists (18.5%) treated two, (3.8%) three, and (1.3%) four youths. Similarly, most (86.8%) youths were treated by one therapist; the rest were treated by two or three therapists (within the same treatment condition) when the initially assigned therapist(s) became unavailable (e.g., left the clinic before treatment was completed.

Youth characteristics. Youths were aged 7 – 15 years (M = 11.42, SD = 2.16), 57.1% female, and represented a range of races/ethnicities and socioeconomic levels. Caucasian American youths and Latino/Latina American youths each comprised 29.7% of the sample, African American youths comprised 17.6%, mixed or other race/ethnicity youths comprised 12.1%, and 11.0% did not report their ethnicity. Median annual family income was in the $15,000 to $30,000 bracket. Primary diagnoses were generated using the Diagnostic Interview Schedule for Children Version 4 (DISC 4.0), combined youth and parent report (Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000—described below). Primary diagnoses were major depressive disorder (29.7%), minor depressive disorder (18.7%), dysthmic disorder (8.8%), separation anxiety disorder (16.5%), social phobia (11.0%), generalized anxiety disorder (6.6%), and other anxiety disorders (8.8%, e.g., agoraphobia, specific phobias). Comorbid disorders included depressive or anxiety disorders other than the primary diagnosis, attention-deficit/hyperactivity disorder, oppositional defiant disorder, and conduct disorder. Youths diagnosed
with a psychotic disorder, pervasive developmental disorder, or mental retardation were excluded.

**Therapist characteristics.** The therapists ranged in age from 25 to 65 years \((M = 32.62, SD = 8.25)\) and 71.3% were female (18.8% male, 10.0% not reported). Caucasian therapists made up 45.0% of the sample, Latino/Latina therapists 25.0%, Asian/Pacific Islander therapists 8.8%, African American therapists 2.5%, mixed or other race/ethnicity therapists 7.5%, and 11.3% of therapists did not report their ethnicity. Two-thirds (62.5%) of the therapists had prior academic training in psychology (comprising 18.0% PhD or PsyD, 78.0% MA, MS, or MEd, 4.0% BA or BS); 22.5% received training in social work (all MSW, LCSW, or both); 6.25% received training in other professional fields (e.g., art therapy, all had MA degree, or a Marriage, Family and Child Counseling license, or both); and 8.75% did not report information on their professional training. On average, therapists had received 4.33 \((SD = 1.77)\) years of professional training and another 3.60 \((SD = 6.36)\) years of postgraduate training experience.

**Treatment Conditions**

CBT therapists conducted Primary and Secondary Control Enhancement Training (PASCET; Weisz, Thurber, Sweeney, Proffitt, & LeGagnoux, 1997) in the depression RCT and Coping Cat (CC; Kendall et al., 1990) in the anxiety RCT. PASCET and CC are manualized treatments designed to last 10 to 15 sessions, and 16 to 20 sessions respectively. In both RCTs, UC therapists used treatment strategies from various orientations that they normally used in their practice and terminated according to their usual practice. All therapists received supervision as usual; in addition, CBT therapists received six hours of training for each manualized treatment and 30 minutes of weekly supervision from doctoral-level psychologists.
Youth- and Parent-Report Measures

Youths and parents completed the following measures of youth internalizing problems and coping at pretreatment, posttreatment, and follow-up. Posttreatment assessment occurred at a mean of 2.58 (SD = 2.31) months after the last treatment session; follow-up assessment occurred at a mean of 16.93 (SD = 7.08) months after the last treatment session. For all measures, higher scores reflect greater symptom severity or more frequent use of that dimension of coping.

Diagnostic Interview Schedule for Children 4. The DISC 4.0 (Shaffer et al, 2000) is a structured diagnostic interview comprising yes/no questions designed for administration by lay interviewers to generate DSM-IV (American Psychiatric Association, 2000) diagnoses. It was used in the present study to generate diagnoses and symptom counts. The present analyses used four continuous symptom count measures: youth-reported depression, parent-reported depression, youth-reported anxiety, and parent-reported anxiety.

Child Behavior Checklist (CBCL). The CBCL (Achenbach, 1991; Achenbach & Rescorla, 2001) is a 118-item parent-report measure of emotional and behavioral problems among youths that is widely used, highly reliable, and well-validated. The T-scores of the broadband Internalizing problems scale (Cronbach’s α = .84) were used in the present analyses.

Children’s Depression Inventory (CDI). The CDI; Kovacs, 2003) is 27-item youth self-report measure of depression symptoms with a parallel parent-report form (CDI-P). The total score of the CDI (Cronbach’s α = .89) and CDI-P (Cronbach’s α = .82) were used in present analyses.

State-Trait Anxiety Inventory for Children-Trait Version (STAIC-T). The STAIC-T (Spielberger, 1973) is a 20-item youth self-report measure of anxiety symptoms with a parallel parent-report form (STAIC-P-T). The total score of the STAIC-T (Cronbach’s α = .85) and
STAIC-P-T (Cronbach’s $\alpha = .85$) were used in present analyses.

**Modified Responses to Stress Questionnaire (MRSQ).** The RSQ (Connor-Smith et al. (2000) is a widely-used and validated measure of coping with parallel youth- and parent-report forms for rating how often the youths used or experienced 57 responses to stress on a four-point scale. There are several versions assessing responses to social stress, to family and economic strain, and to recurrent abdominal pain (Connor-Smith et al., 2000). Participants in the present study received one of two versions of the MRSQ: youths with a primary diagnosis of a depressive disorder and their parents received the version that asks about youth coping when feeling *sad* or *bad*, and youths with a primary diagnosis of an anxiety disorder and their parents received the version asks about coping when feeling *scared* or *worried*. Aside from this difference in instructions, items (i.e., the responses to stress being rated) were identical across the *sad* and *scared* versions. The MRSQ has 54 items because two RSQ items that were not appropriate for the MRSQ given the modification, and one RSQ item that was similar to another item, were excluded from the MRSQ. The items are grouped into subscales comprising three items, three to five subscales are in turn categorized under each of five dimensions of responses to stress: Primary Control Engagement Coping, Secondary Control Engagement Coping, Disengagement Coping, Involuntary Engagement, and Involuntary Disengagement. The dimension scores were computed by first taking the mean rating of items in each subscale, and then the mean rating of subscales under each dimension; next, each dimension score was divided by the sum of the five dimension scores to give a coping proportion score (see Connor-Smith et al., 2000; Jaser et al., 2005). The proportion scores have the advantage of being corrected for differences in response bias among participants (i.e. some participants may rate most methods of coping as high in frequency, others may rate most methods of coping as low in frequency). Four
scores were computed for the present analyses, youth-reported primary control (Cronbach’s $\alpha = .80$), parent-reported primary control (Cronbach’s $\alpha = .81$), youth-reported secondary control (Cronbach’s $\alpha = .81$), and parent-reported secondary control (Cronbach’s $\alpha = .75$). Correlations between parent- and youth-report ranged from -.01 to .32 (mean $r = .22$) within coping dimension and time point, comparable to the correlations between primary and secondary control scores within informant and time point (mean $r = .29$, range = .02 to .47), and consistent with prior findings on the level of association between parent and youth reports (e.g., Achenbach, McConaughy & Howell, 1987). Thus coping scores were kept as separate measures for each informant.

**Observational Coding Measures**

Treatment sessions were recorded and sampled for observational coding of therapeutic alliance and therapists’ use of different strategies. Coding was carried out using two scales of the Therapy Process Observational Coding System for Child Psychotherapy (TPOCS; McLeod & Weisz, 2005, 2010), detailed below. Higher scores on the scales reflect stronger alliance or more extensive use of strategies for that treatment orientation.

**Therapy Process Observational Coding System for Child Psychotherapy – Alliance scale (TPOCS-A).** Therapeutic alliance was coded using the youth–therapist alliance form of the TPOCS–A (McLeod & Weisz, 2005). This form comprises nine items on the client-therapist bond and client participation in therapy tasks, each rated on a six-point scale. Ratings of the parent–therapist alliance were available for only a small subsample of participants and thus were not included in the present analysis. Because the number of sessions varied substantially across youths ($M = 18.26$, $SD = 10.87$, range 2 – 65), therapy sessions were sampled for coding by first dividing each youth’s treatment course into three phases, and then randomly selecting one
session from the beginning phase excluding the first session, two sessions from the middle phase, and one session from the end phase excluding the last session (Langer, McLeod, & Weisz, 2011; McLeod & Weisz, 2005). When four recorded sessions were not available, all sessions were coded. When recorded sessions were available for multiple therapists per youth, sessions were selected from each therapist. For the present analyses, an early alliance score was computed by taking the mean of the two earliest sessions in the first half of treatment including the median session and excluding any sessions after session 9; a late alliance score was computed by taking the mean of the two latest sessions in the second half of treatment. This approach of averaging scores across two early and two late sessions (when available) has been taken by other researchers to enhance the reliability of alliance coding (e.g., Chiu et al., 2009). The decision to limit early alliance to sessions through session 9 was informed by empirical findings that symptoms did not impact alliance through session 9 of a 16-session course of therapy (Crits-Christoph, Gibbons, Hamilton, Ring-Kurtz, & Gallop, 2011)—consistent with the purpose of computing an early alliance score that is unlikely to be influenced by symptom improvement.

These procedures generated a sample of 258 coded sessions for the present analyses, all double-coded (intercoder ICC = .80, Cronbach’s α for items = .92) by four post-Bachelor’s degree research assistants, four graduate students, and a doctoral-level licensed psychologist blind to treatment condition and outcomes. Scores for each session were averaged across coders before computing early alliance and late alliance scores.

**Therapy Process Observational Coding System for Child Psychotherapy – Strategies scale (TPOCS-S).** The TPOCS-S (McLeod & Weisz, 2010) comprises intervention strategies consistent with cognitive, behavioral, psychodynamic, family systems, and client-centered orientations. A seven-point scale was used to rate the thoroughness and frequency with
which the therapist used each strategy. The present analyses drew from a larger pool of sessions coded for the validation study (McLeod & Weisz, 2010) of this measure and the primary outcome studies (Southam-Gerow et al., 2010; Weisz et al., 2009). Given the slightly different sampling procedures used in the validation and outcome studies, the following approach was applied for the present analyses: two sessions were randomly selected for treatment duration up to 20 sessions, three sessions for duration up to 40 sessions, and four sessions for duration over 40 sessions, excluding the first session. At least one session each was selected from the first and second halves of therapy, and two sessions were selected from each of multiple therapists, whenever available. These procedures generated a sample of 160 coded sessions for the present analyses. Strategy scores were computed in a similar manner to the coping dimension proportion scores—by taking the mean rating of the items within each orientation and dividing it by the sum of the five mean orientation scores. Scores over multiple sessions were averaged to form a single TPOCS-S score for each strategy type per youth, among which the cognitive strategy (Cronbach’s α = .77) and behavioral strategy (Cronbach’s α = .76) and scores were used in the present analyses. Coders were three graduate students, one of whom designed the coding manual and served as the master coder. To establish intercoder reliability, two samples of 30 sessions were independently coded by the master coder and one other coder (cognitive score mean ICC = .80, behavioral score mean ICC = .85). The remaining sessions were single-coded by each of the three coders. The present analyses utilized the mean of the double-coded scores for the reliability sample, and single-coded scores for the rest of the sample.

**Data Analytic Plan**

**Missing data analysis.** All outcome measures had at least 75% of items completed; missing item-level data were handled by computing a prorated total score (i.e., equivalent to
assuming that missing items were given the participant’s average rating of the completed measure). Missing items on the MRSQ were also handled following scoring instructions for this measure; specifically, a subscale score was computed from the mean of item ratings if at least two of three items were available, and a dimension score was computed from the mean of subscale scores if at least two of three to five dimension scores were available. No item-level data were missing on the TPOCS-A and TPOCS-S. After taking these steps, the following data remained missing at the measure level: 2.5% of youth-reported and 2.4% of parent-reported pretreatment outcomes, 11.0% of youth-reported and 10.3% parent-reported posttreatment outcomes, and 40.1% of youth-reported and 40.4% parent-reported follow-up outcomes, averaged across different outcome measures. MRSQ scores were missing for 22.0% of youth-reported and 20.9% of parent-reported measures at pretreatment, 13.2% of youth-reported and 15.4% of parent-reported measures at posttreatment, and 40.7% of both youth-and parent-reported measures at follow-up. There was a greater proportion of MRSQ scores missing at pretreatment than at posttreatment because the measure was added to the study protocol after recruitment began. Due to the high attrition rate at follow-up, follow-up data were not used in the mediation analyses to answer research questions—nevertheless, they were retained to facilitate the analysis and imputation of missing data. On the TPOCS-S, early alliance, late alliance scores, and strategy scores were missing for 15.4%, 19.8%, and 23.1% of youths respectively. SPSS Version 24 Missing Value Analysis (MVA) was employed to examine missing data patterns for variables with more than 5% missing data at the measure level. Separate variance $t$-tests were generated to show whether missingness among key variables to be analyzed in the present study (i.e., pre- and posttreatment outcomes, pre- and posttreatment primary control and secondary control coping, alliance, and therapist cognitive and behavioral strategies) were significantly
predicted by one another, by study design variables (e.g., treatment condition, primary diagnosis), or by auxiliary variables that will not be used in mediation analyses (e.g., demographics, treatment duration, follow-up data, other responses to stress, other therapist strategies). Familywise Type 1 error rate was used for each variable. For example, 72 t-tests were conducted to examine whether missingness on early alliance scores were predicted by 72 other variables, giving $p < .0007 (0.05/72)$ as the familywise Type 1 error. Missing early alliance scores were significantly predicted by higher pretreatment youth-reported depression symptom counts ($p = .0005$). Missing posttreatment CDI scores were significantly predicted by longer time between the last treatment session and posttreatment assessment ($p = .0001$). Missing posttreatment youth-reported depression symptom counts were predicted by lower posttreatment parent-reported depression symptom counts ($p < .0001$). Missing posttreatment parent-reported anxiety and depression symptom counts were both predicted by lower posttreatment youth-reported anxiety symptom counts ($p < .0001$), and by higher follow-up STAIC-P-T scores ($p = .0008$). To summarize, missingness on the predictor and mediator variables for the planned analyses were not predicted by any posttreatment outcome measure. However, missingness on some posttreatment outcome measures was predicted by other posttreatment outcome measures (all DISC symptom counts), suggesting that missingness on those outcomes may be missing not at random (MNAR; see Tabachnick & Fidell, 2007). To mitigate any potential distortions of the variable relationships to be analyzed in the present study due to MNAR, multiple imputation (MI) was first performed on the measure-level variables. Next, composites were created from individual outcome measures (with imputed data) by informant, so that bias due to missingness on any one outcome could be countered by other outcomes from the same informant.
Multiple imputation. Multiple imputation is considered the gold standard technique for addressing missing data, and it does not assume that data are missing at random (MAR; see Tabachnick & Fidell, 2007). Briefly, it involves using the distribution of available data to estimate multiple sets of plausible values for the missing data, with the addition of random components into the estimated values; the same analysis is conducted on the multiple datasets with imputed values, and the results are combined across datasets in a way that accounts for the uncertainty in the results due to missing data within and between datasets (White, Royston, & Wood, 2011). For the present study, several recommendations by White et al. (2011) were followed to optimize the multiple imputation procedure. First, given that the highest percentage of incomplete cases among the planned mediation analyses is 47.3%, 50 imputed datasets were utilized based on the rule of thumb that reproducible results requires at least as many datasets as the percentage of incomplete cases. Second, follow-up and other auxiliary data that would not be used in the mediation analyses, but that might predict values or missingness on the mediation model variables, were included to make the assumption of MAR more likely, reduce bias, and improve the accuracy and precision of the imputed estimates. Third, to increase the power to detect interactions in the mediation analyses between treatment condition, primary diagnosis, and the continuous mediator/predictor variables (coping, alliance, and therapist strategies), interaction terms were computed between primary diagnosis and each continuous mediator/predictor using available data, and then the dataset was stratified by treatment condition and imputation was conducted separately for each condition. Multiple imputation was implemented using IVEware (Raghunathan, Lepkowski, Van Hoewyk, & Solenberger, 2001; Raghunathan, Solenberger, & Van Hoewyk, 2002) via SAS.

Compositing internalizing symptom outcomes. The five parent-report outcome
measures (DISC depression and anxiety symptom counts, CBCL, CDI-P, STAIC-P-T) were
moderately to strongly correlated with one another at each time point (mean $r = .54$, range .35 – .79). Correlations among the four youth-report outcome measures (DISC depression and anxiety symptom counts, CDI, STAIC-T) were similar to those found among parent-report measures except for the correlation ($r = .17$) between CDI score and DISC anxiety symptom count at pretreatment (mean $r = .59$, range of other correlations .30 – .76). Correlations between informants across various measures, but within time point, were considerably smaller (mean $r = .17$, range -.04 – .37). This pattern supports computing a composite outcome by informant using the unit weighting described in Weisz et al. (2009) and Southam-Gerow et al. (2010).

Specifically, the pretreatment scores of the five parent-report symptom measures were standardized and then averaged for each participant to generate a pretreatment parent-reported composite internalizing symptom outcome within each imputed dataset. To generate a posttreatment parent-reported composite internalizing outcome comparable to the pretreatment composite, the posttreatment scores of the five parent-report measures were standardized based on the means and standard deviations of the pretreatment scores and averaged for each participant. The same procedure was performed with the four youth-report symptom measures to generate pretreatment and posttreatment youth-reported composite internalizing symptom outcomes, and then repeated in all 50 imputed datasets

**Descriptive analyses.** Descriptive analyses of coping, alliance, outcomes, and therapist strategies were conducted to provide context for interpreting the results of the mediation models that are the focus of the present study. Pre–post change scores were computed for all coping and outcome variables, and early–late change was computed for alliance. The mediation analyses only indicate whether any potential change differs across levels of the predictor, whereas
analysis of change scores indicate whether the mediator and outcome variables change across time and in what direction. One sample \( t \)-tests were conducted using SAS PROC UNIVARIATE to determine if the pre–post changes were significantly different from zero for the sample as a whole. The change scores and the therapist cognitive strategy and behavioral strategy scores were each regressed on treatment condition to test for differences between MCBT and UC using SAS PROC REGRESSION. The mean of parameter estimates across the 50 imputed datasets was taken, and the standard errors (SEs) were combined using Rubin’s (1987) rules and used to generate \( p \)-values, as implemented by SAS PROC MIANALYZE.

**Mediation models.** To answer the primary research question of whether each candidate mechanism mediates treatment outcomes, simple mediation models were fitted to assess posttreatment primary control coping, secondary control coping and late alliance separately as mediators of the effect of treatment condition on posttreatment internalizing symptoms, controlling for pretreatment coping or early alliance, pretreatment internalizing symptoms, and primary diagnosis. This first set of analyses was limited to six models by matching each of the four coping measures to the outcome derived from the same informant, and alliance to both youth-and parent-reported outcomes. A second set of analyses, identical except for the addition of the interaction of treatment condition with the candidate mechanism as a predictor, was conducted following Kraemer and colleagues’ (2002) recommendation, the rationale being that treatment may moderate the effect of the mediator on the outcome (i.e., the relationship between the mediator and outcome may be different in the two treatment conditions). This second type of mediation is a specific type of *moderated mediation* or *conditional indirect effect* in which the indirect effect depends on the treatment condition; there can also be *conditional direct effects* that depend on the level of the mediator variable (Hayes, 2013; Preacher, Rucker, & Hayes,
2007). Figures 2.1 and 2.2 depict the simple mediation model and the moderation mediation model respectively, along with the paths used to compute indirect and direct effects in each model type. To investigate whether the treatment effects being mediated were in fact therapists’ use of cognitive strategies or behavioral strategies, regardless of whether they were employed in the context of manualized CBT or usual care, the model(s) demonstrating significant mediation would be modified by replacing treatment condition with therapist cognitive or behavioral strategies as the predictor. If the effects of therapist cognitive or behavioral strategies were in fact mediated by one or more candidate mechanisms, the model(s) would be further modified to examine whether the mediation effects are moderated by treatment condition, primary diagnosis, or both.
Figure 2.1. In the simple mediation model, mediation occurs when the predictor is associated with the mediator variable (path $a$) and the mediator variable is associated with the outcome (path $b$). Thus the predictor can be said to influence the outcome indirectly through the mediator variable. The mediation or indirect effect is computed by taking the product of the regression coefficients of the two paths, $ab$, and is significant when the 95% confidence interval (CI) of $ab$ does not contain zero. The direct effect is the association between the predictor and the outcome that is not accounted for by the mediator variable, $c'$ (i.e., the association between the predictor and the outcome controlling for effects of the mediator on the outcome). The direct effect is significant when the 95% CI of $c'$ does not contain zero.
Figure 2.2. In this moderated mediation model, mediation occurs when the predictor is associated with the mediator variable (path $a$), and: either the mediator variable is associated with the outcome (path $b$), or the interaction of the predictor $\times$ mediator variable is associated with the outcome (path $c'_{2}$). Thus the mediation or indirect effect depends in part on the level of the predictor ($X$). The mediation effect is the sum of: the product of the regression coefficient of the first two paths, $ab$, and the product of the regression coefficients of the first and third path and the level of the predictor, $ac_{2}'X$. Mediation is significant at a particular level of the predictor when the 95% confidence interval (CI) of $ab + ac_{2}'X$ does not contain zero. Moderated mediation occurs when indirect effects are significantly different at two different levels of the predictor, or when the 95% CI of $ac_{2}'$ (i.e., the portion of the indirect effect that is conditional on $X$) does not contain zero. The direct effect can also be moderated; it depends in part on the level of the mediator ($M$). The direct effect is the sum of: the association between the predictor and the outcome that is not accounted for by the mediator variable, $c_{1}'$, and the product of the predictor $\times$ mediator variable interaction and the level of the mediator, $c_{2}'M$. The direct effect is significant at a particular level of the mediator variable when the 95% CI of $c_{1}' + c_{2}'M$ does not contain zero. The direct effect is moderated when direct effects are significantly different at two different levels of the mediator variable, or when the 95% CI of $c_{2}'$ (i.e., the portion of the direct effect that is conditional on $M$) does not contain zero. Figure adapted from Hayes (2013).
Combining bootstrapping and multiple imputation. A bootstrapping approach (Preacher & Hayes, 2008) was employed to test for simple and moderated mediation effects because it offers several advantages: it does not require treatment group differences in symptom outcome measures as a precondition, it is more highly powered than other approaches, and it does not require the strict assumption that the sampling distribution of the indirect effect be normal—even if the parameter estimates \( a \) and \( b \) are each normally distributed, their product \( ab \) is unlikely to be normally distributed (Preacher & Hayes, 2008). Bootstrapping involves creating a new sample by randomly selecting a data point from the original study sample and
replacing it with an identical data point in the original sample, then randomly selecting another data point before replacing it in the original sample, so that the identical replacements of previously selected data points can be selected any number of times. A number of data points equivalent to the original sample size is selected, generating a new bootstrapped sample the same size as the original sample, and comprising most likely a subset of data points from the original study sample, plus some identical replacements of data points that were selected. This process is repeated to generate many bootstrapped samples that vary in the specific subset of data points selected and the number of identical replacements per selected data point. The indirect effect is computed from every bootstrapped sample, forming an empirical sampling distribution (ESD) of the indirect effect. The mean of this ESD is the estimate of the true indirect effect. A 95% confidence (CI) around this mean that does not include zero indicates that the indirect effect is significantly different from zero and that mediation effects were significant. Thus bootstrapping is increasingly popular for testing mediation effects with single observed datasets, but bootstrapping with multiply imputed datasets is still an active area of research (Wu & Jia, 2013). The present study used a modified version of the MI(BOOT) procedure (i.e., bootstrapping nested within MI) recommended by Wu and Jia (2013) to obtain bias corrected bootstrapped CIs of the indirect effect for each mediation model across the 50 imputed datasets. First, the PROCESS macro in SAS version 9.4 (Hayes, 2013) was used to fit the mediation model to each of the 50 imputed datasets, generating 50 point estimates of indirect effects, direct effects, and other parameter estimates of paths in the mediation model, as well as 5000 bootstrapped samples from each imputed dataset. Second, parameter estimates of model paths for the 5000 bootstrapped samples were outputted and merged across the 50 imputed datasets to produce a dataset of containing 250,000 bootstrapped samples of each parameter estimate. Because
parameter estimates of indirect effects involve combinations of model paths (e.g., \( ab, ab + ac_2'X \), see Figure 2.1) and thus were not automatically outputted, they were computed for the 250,000 bootstrapped samples. The 250,000 bootstrapped samples form a conditional ESD, with each set of 5000 samples conditional on the source imputed dataset (Wu & Jia, 2013). Third, the overall point estimates were obtained by taking the mean of the point estimates across the 50 imputed datasets. Fourth, overall 95% bias corrected CIs were obtained by applying Wu and Jia’s BCCI macro in SAS, which uses the 250,000 bootstrapped samples to construct a 95% CI around the point estimate while adjusting for the difference between the point estimate and median of the conditional ESD due to asymmetry of the ESD (i.e., bias correction). This approach can in fact be used for any parameter estimate. However, due to the laborious and computationally intensive nature of this approach, bootstrapping was used only for indirect effects and direct effects because both simple and conditional indirect effects, as well as conditional direct effects, involve combinations of model paths (see Figures 2.1 and 2.2) and thus may not be normally distributed. Overall individual model paths estimates and SEs were combined using a more efficient procedure: parameter estimates were computed for every imputed dataset via SAS PROC REGRESSION; then the mean of the parameter estimates was taken, and the combination of SEs using Rubin’s (1987) rules was used to generate \( p \)-values, as implemented by SAS PROC MIANALYZE.

**Results**

**Descriptive Analyses**

For the overall sample, parent-reported primary control coping, parent-reported secondary control coping, and youth-reported secondary control coping all increased significantly from pre- to posttreatment (all \( ps < .01 \)). Both youth- and parent-reported
internalizing symptoms decreased significantly from pre- to posttreatment (all $ps < .0001$). Pre– post change for youth-reported primary control coping and early–late change in alliance were not significant ($ps > .05$). When examining change scores by treatment condition, a significant change for youths receiving MCBT was accompanied by significant change for youths receiving UC for all variables except one—youth-reported secondary control increased significantly in UC ($p < .05$) and near significantly in MCBT ($p < .10$). However, the between-treatment differences on the change scores for all variables were nonsignificant. Therapist cognitive and behavioral strategies were used significantly more frequently in MCBT than in UC (all $ps < .001$), as expected. It is interesting to note that the variation in use of cognitive and behavioral strategies among MCBT therapists is about the same as, or greater than, the variation among UC therapists. One might assume that there would be less variation among MCBT therapists in use of cognitive and behavioral strategies as treatment was guided by a structured protocol. Table 2.1 displays the descriptive statistics.

**Mediation Models with Treatment Condition as Predictor**

This section describes the main results of the six simple mediation models and six moderated mediation models with treatment condition as the predictor. Estimates of the treatment effect on the mediator (path $a$), and the total effect of treatment on outcome without the mediator in the model (path $c$), are identical in the simple and moderated mediation models, and are thus only reported once for each candidate mediator. Table 2.2 displays all point estimates, standard errors, and $p$-values of the models.
Table 2.1. *Descriptive Statistics of Coping, Alliance, Outcome, and Therapist Strategy Variables Combined Across 50 Imputed Datasets (N = 91)*

<table>
<thead>
<tr>
<th></th>
<th>Manualized Cognitive Behavioral Therapy (n = 47)</th>
<th>Usual Care (n = 44)</th>
<th>Between-Treatment Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre/Early  Post/Late  Change Score/Score</td>
<td>Pre/Early  Post/Late  Change Score/Score</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M     SE       M     SE       M     SE       M     SE       M     SE       P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth-Reported Primary</td>
<td>0.231 0.009   0.235 0.007   0.003 0.010</td>
<td>0.230 0.007   0.237 0.009   0.007 0.009</td>
<td>.793</td>
</tr>
<tr>
<td>Control Coping</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent-Reported Primary</td>
<td>0.203 0.008   0.242 0.007   0.039*** 0.009</td>
<td>0.222 0.009   0.246 0.008   0.024* 0.011</td>
<td>.300</td>
</tr>
<tr>
<td>Control Coping</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth-Reported Secondary</td>
<td>0.212 0.005   0.227 0.007   0.014† 0.008</td>
<td>0.225 0.009   0.253 0.010   0.028* 0.011</td>
<td>.308</td>
</tr>
<tr>
<td>Control Coping</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent-Reported Secondary</td>
<td>0.182 0.008   0.204 0.006   0.022* 0.009</td>
<td>0.176 0.006   0.217 0.006   0.041*** 0.008</td>
<td>.111</td>
</tr>
<tr>
<td>Control Coping</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth-Therapist Alliance</td>
<td>3.584 0.089   3.489 0.104   -0.095 0.096</td>
<td>3.268 0.115   3.333 0.140   0.065 0.123</td>
<td>.312</td>
</tr>
<tr>
<td>Youth-Reported Symptoms</td>
<td>-0.005 0.105  -0.491 0.139  -0.486** 0.130</td>
<td>0.006 0.119   -0.725 0.136  -0.731*** 0.110</td>
<td>.155</td>
</tr>
<tr>
<td>Parent-Reported Symptoms</td>
<td>-0.018 0.120  -0.827 0.131  -0.809*** 0.114</td>
<td>0.019 0.114   -0.917 0.135  -0.936*** 0.125</td>
<td>.447</td>
</tr>
<tr>
<td>Therapist Cognitive</td>
<td>–     –       –     –       0.032 0.009</td>
<td>–     –       –     –       -0.038 0.010</td>
<td>.001</td>
</tr>
<tr>
<td>Strategies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Therapist Behavioral</td>
<td>–     –       –     –       0.035 0.014</td>
<td>–     –       –     –       -0.034 0.007</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Strategies</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* M = mean, SE = standard error. †p < .10; *p < .05; **p < .01, ***p < .001. Therapist cognitive and behavioral strategy scores were centered.
Table 2.2. Regression Models Fitted to Assess Youth Coping and Therapeutic Alliance as Simple Mediators and Moderated Mediators of the Effect of Treatment Condition on Posttreatment Internalizing Symptoms, Covarying Pretreatment Symptoms, Pretreatment Coping or Early Alliance, and Primary Diagnosis, With Parameters Combined Across 50 Imputed Datasets (N = 91)

<table>
<thead>
<tr>
<th>Predictor/Covariate</th>
<th>Regression Model Predicting Mediator Variable</th>
<th>Regression Models Predicting Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mediator Variable: Posttreatment Youth-Reported Primary Control Coping</td>
<td>Outcome: Posttreatment Youth-Reported Symptoms</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.169</td>
<td>0.396</td>
</tr>
<tr>
<td></td>
<td>0.042</td>
<td>0.574</td>
</tr>
<tr>
<td></td>
<td>&lt;.001</td>
<td>.490</td>
</tr>
<tr>
<td>Pretreatment Youth-Reported Symptoms</td>
<td>-0.016</td>
<td>0.595</td>
</tr>
<tr>
<td></td>
<td>0.007</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>.027</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Pretreatment Youth-Reported Primary Control Coping</td>
<td>0.317</td>
<td>-1.066</td>
</tr>
<tr>
<td></td>
<td>0.171</td>
<td>2.247</td>
</tr>
<tr>
<td></td>
<td>.066</td>
<td>.636</td>
</tr>
<tr>
<td>Primary Diagnosis (depression=0, anxiety=1)</td>
<td>-0.009</td>
<td>-0.109</td>
</tr>
<tr>
<td></td>
<td>0.011</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td>.435</td>
<td>.468</td>
</tr>
<tr>
<td>Treatment Condition (UC=0, MCBT=1)</td>
<td>-0.004&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.222&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>0.011</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>.721</td>
<td>.176</td>
</tr>
<tr>
<td>Posttreatment Youth-Reported Primary Control Coping</td>
<td>–</td>
<td>-3.488&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Treatment Condition × Posttreatment Youth-</td>
<td>–</td>
<td>1.875</td>
</tr>
<tr>
<td>Reported Primary Control Coping</td>
<td>–</td>
<td>.064</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
Table 2.2 (Continued).

<table>
<thead>
<tr>
<th></th>
<th>Mediator Variable: Posttreatment Parent-Reported Primary Control Coping</th>
<th>Outcome: Posttreatment Parent-Reported Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simple Mediation Model</td>
<td>Moderated Mediation Model</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.178 ± 0.030</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Pretreatment Parent-Reported Symptoms</td>
<td>0.001 ± 0.007</td>
<td>.912</td>
</tr>
<tr>
<td>Pretreatment Parent-Reported Primary Control Coping</td>
<td>0.300 ± 0.131</td>
<td>.022</td>
</tr>
<tr>
<td>Primary Diagnosis (depression=0, anxiety=1)</td>
<td>0.003 ± 0.010</td>
<td>.808</td>
</tr>
<tr>
<td>Treatment Condition (UC=0, MCBT=1)</td>
<td>0.002 ± 0.011</td>
<td>.878</td>
</tr>
<tr>
<td>Posttreatment Parent-Reported Primary Control Coping</td>
<td>- ± -</td>
<td>-</td>
</tr>
<tr>
<td>Treatment Condition × Posttreatment Parent-Reported Primary Control Coping</td>
<td>- ± -</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mediator Variable: Posttreatment Youth-Reported Secondary Control Coping</th>
<th>Outcome: Posttreatment Youth-Reported Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simple Mediation Model</td>
<td>Moderated Mediation Model</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.170 ± 0.037</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Pretreatment Youth-Reported Symptoms</td>
<td>-0.023 ± 0.008</td>
<td>.004</td>
</tr>
<tr>
<td>Pretreatment Youth-Reported Secondary Control Coping</td>
<td>0.374 ± 0.162</td>
<td>.021</td>
</tr>
<tr>
<td>Primary Diagnosis (depression=0, anxiety=1)</td>
<td>-0.002 ± 0.011</td>
<td>.829</td>
</tr>
</tbody>
</table>
Table 2.2 (Continued).

<table>
<thead>
<tr>
<th>Treatment Condition (UC=0, MCBT=1)</th>
<th>-0.022(^a)</th>
<th>0.011</th>
<th>.055</th>
<th>0.095(^c)</th>
<th>0.163</th>
<th>.563</th>
<th>1.217(^b)</th>
<th>0.719</th>
<th>0.091</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttreatment Youth-Reported Secondary Control Coping</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-5.755(^b)</td>
<td>1.699</td>
<td>.001</td>
<td>-3.964</td>
<td>2.051</td>
<td>0.054</td>
</tr>
<tr>
<td>Treatment Condition × Posttreatment Youth-Reported Secondary Control Coping</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-4.764(^c_2)</td>
<td>2.973</td>
<td>0.109</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mediator Variable: Posttreatment Parent-Reported Secondary Control Coping</th>
<th>Outcome: Posttreatment Parent-Reported Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.195</td>
</tr>
<tr>
<td>Pretreatment Parent-reported Symptoms</td>
<td>-0.012</td>
</tr>
<tr>
<td>Pretreatment Parent-reported Secondary Control Coping</td>
<td>0.115</td>
</tr>
<tr>
<td>Primary Diagnosis (depression=0, anxiety=1)</td>
<td>0.003</td>
</tr>
<tr>
<td>Treatment Condition (UC=0, MCBT=1)</td>
<td>-0.014(^a)</td>
</tr>
<tr>
<td>Posttreatment Parent-Reported Secondary Control Coping</td>
<td>-</td>
</tr>
<tr>
<td>Treatment Condition × Posttreatment Parent-Reported Secondary Control Coping</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 2.2 (Continued).

<table>
<thead>
<tr>
<th></th>
<th>Mediator Variable:</th>
<th>Outcome: Posttreatment Youth-Reported Symptoms</th>
<th>Outcome: Posttreatment Parent-Reported Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Late Alliance</td>
<td>Simple Mediation Model</td>
<td>Moderated Mediation Model</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.210</td>
<td>-0.259</td>
<td>-0.697</td>
</tr>
<tr>
<td>Pretreatment Youth-Reported Symptoms</td>
<td>0.033</td>
<td>0.693</td>
<td>0.702</td>
</tr>
<tr>
<td>Early Alliance</td>
<td>0.657</td>
<td>0.032</td>
<td>0.032</td>
</tr>
<tr>
<td>Primary Diagnosis (depression=0, anxiety=1)</td>
<td>-0.053</td>
<td>-0.061</td>
<td>-0.088</td>
</tr>
<tr>
<td>Treatment Condition (UC=0, MCBT=1)</td>
<td>-0.056^a</td>
<td>-0.164^c</td>
<td>1.413^c_1</td>
</tr>
<tr>
<td>Late Alliance</td>
<td>–</td>
<td>0.252^b</td>
<td>-0.028^b</td>
</tr>
<tr>
<td>Treatment Condition × Late Alliance</td>
<td>–</td>
<td>–</td>
<td>-0.340^c_2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>1.198</td>
<td>-0.513</td>
<td>-0.592</td>
</tr>
<tr>
<td>Pretreatment Parent-reported Symptoms</td>
<td>0.010</td>
<td>0.600</td>
<td>0.601</td>
</tr>
<tr>
<td>Early Alliance</td>
<td>0.661</td>
<td>0.086</td>
<td>0.086</td>
</tr>
<tr>
<td>Primary Diagnosis (depression=0, anxiety=1)</td>
<td>-0.052</td>
<td>-0.267</td>
<td>-0.271</td>
</tr>
<tr>
<td>Treatment Condition (UC=0, MCBT=1)</td>
<td>-0.058^a</td>
<td>0.087^c</td>
<td>0.300^c_1</td>
</tr>
<tr>
<td>Late Alliance</td>
<td>–</td>
<td>-0.171^b</td>
<td>-0.147^b</td>
</tr>
<tr>
<td>Treatment Condition × Late Alliance</td>
<td>–</td>
<td>–</td>
<td>-0.062^c_2</td>
</tr>
</tbody>
</table>
Table 2.2 (Continued).

Note. $B =$ unstandardized regression coefficient, $SE =$ standard error, UC=usual care, MCBT=manualized cognitive behavioral therapy.

The regression models predicting each mediator variable are identical for simple and moderated mediation models and are thus displayed only once per mediator variable.

$^a$Denotes regression coefficient for path $a$ of mediation model. $^b$Denotes regression coefficient for path $b$ of mediation model.

$^c$Denotes regression coefficient for path $c'$ of simple mediation model. $^c_1$Denotes regression coefficient for path $c_1'$ of moderated mediation model. $^c_2$Denotes regression coefficient for path $c_2'$ of moderated mediation model.
Youth-reported primary control coping. In the simple mediation model, the indirect effect of treatment condition on outcome via youth-reported primary control coping was not significant ($B = 0.012, \text{CIs } [-0.056, 0.132]$). Treatment condition had no significant effect on posttreatment coping ($B = -0.004, p = .721)$, but more frequent posttreatment coping had a near significant association with a lower level of symptoms at posttreatment ($B = -3.488, p = .064$), controlling for pretreatment coping and symptoms as well as primary diagnosis. Neither the direct effect ($B = 0.222, \text{CIs } [-0.096, 0.553]$) nor the total effect of treatment condition ($B = 0.234, p = .156$) were significant. In the moderated mediation model, the indirect effects for MCBT ($B = 0.022, \text{CIs } [-0.084, 0.229]$) and UC ($B = 0.007, \text{CIs } [-0.043, 0.112]$) were not significantly different from zero or from each other ($B = 0.015, \text{CIs } [-0.050, 0.225]$). The treatment condition $\times$ posttreatment coping interaction ($B = -3.564, p = .268$) added in the moderated mediation model was also not significantly associated with posttreatment symptom severity. The direct effects of treatment condition on outcome at high levels of posttreatment primary control coping ($B = 0.044, \text{CIs } [-0.406, 0.461]$) and at low levels of posttreatment primary control coping ($B = 0.400, \text{CIs } [-0.124, 0.912]$) were also not significantly different from zero or from each other ($B = -0.356, \text{CIs } [-1.052, 0.353]$). These findings indicate that youth-reported primary control coping is neither a predictor nor mediator of treatment outcome.

Parent-reported primary control coping. In the simple mediation model, the indirect effect of treatment condition on outcome via parent-reported primary control coping was not significant ($B = -0.011, \text{CIs } [-0.163, 0.113]$). Treatment condition had no significant effect on posttreatment coping ($B = 0.002, p = .878$), but more frequent posttreatment coping was significantly associated with a lower level of symptoms ($B = -5.779, p = .004$), controlling for pretreatment coping and symptoms as well as primary diagnosis. Neither the direct effect ($B =
0.094, CIs [-0.234, 0.388]) nor total effect of treatment ($B = 0.083, p = .604$) were nonsignificant. In the moderated mediation model, the indirect effects for MCBT ($B = -0.007$, CIs [-0.163, 0.129]) and UC ($B = -0.015$, CIs [-0.195, 0.105]) were not significantly different from zero or from each other ($B = 0.008$, CIs [-0.038, 0.163]). The treatment condition × posttreatment coping interaction ($B = 0.478, p = .896$) added in the moderated mediation model was not significantly associated with posttreatment symptom severity. The direct effects of treatment condition on outcome at high levels of posttreatment primary control coping ($B = 0.116$, CIs [-0.389, 0.591]) and at low levels of posttreatment primary control coping ($B = 0.074$, CIs [-0.315, 0.507]) were also not significantly different from zero or from each other ($B = 0.043$, CIs [-0.645, 0.668]). These findings indicate that parent-reported primary control coping is a predictor but not a mediator of treatment outcome.

**Youth-reported secondary control coping.** In the simple mediation model, the indirect effect of treatment condition on outcome via youth-reported secondary control coping was significant ($B = 0.125$, CIs [0.002, 0.319]). UC led to near significantly more frequent posttreatment secondary coping than MCBT ($B = -0.022, p = .055$), which was significantly associated with a lower level of symptoms at posttreatment ($B = -5.755, p < .001$), controlling for pretreatment coping and symptoms as well as primary diagnosis. Neither the direct effect ($B = 0.095$, CIs [-0.214, 0.425]) nor total effect of treatment ($B = 0.219, p = .189$) were significant. In the moderated mediation model, the indirect effect for MCBT ($B = 0.192$, CIs [0.0005, 0.530]) was significant but that for UC ($B = 0.008$, CIs [-0.016, 0.261]) was not; however, the two indirect effects were not significantly different from each other ($B = 0.108$, CIs [-0.006, 0.479]). The treatment condition × posttreatment coping interaction ($B = -4.764, p = .109$) added in the moderated mediation model was also not significantly associated with posttreatment symptom
severity. The direct effects of treatment condition on outcome at high levels of posttreatment secondary control coping ($B = -0.182, \text{CIs} [-0.664, 0.283]$) and at low levels of posttreatment secondary control coping ($B = 0.337, \text{CIs} [-0.114, 0.790]$) were also not significantly different from zero or from each other ($B = -0.520, \text{CIs} [-1.234, 0.115]$). These findings indicate that youth-reported secondary control coping is a mediator of treatment outcome.

**Parent-reported secondary control coping.** In the simple mediation model, the indirect effect of treatment condition on outcome via parent-reported secondary control coping was not significant ($B = 0.079, \text{CIs} [-0.010, 0.239]$). Treatment condition had no significant effect on posttreatment coping ($B = -0.014, p = .121$), but more frequent coping was significantly associated with a lower level of symptoms at posttreatment ($B = -5.761, p = .005$), controlling for pretreatment coping and symptoms as well as primary diagnosis. Neither the direct effect ($B = -0.010, \text{CIs} [-0.289, 0.288]$) nor total effect of treatment ($B = 0.069, p = .659$) were nonsignificant. In the moderated mediation model, the indirect effects for MCBT ($B = 0.072, \text{CIs} [-0.012, 0.247]$) and UC ($B = 0.089, \text{CIs} [-0.010, 0.318]$) were not significantly different from zero or from each other ($B = -0.017, \text{CIs} [-0.218, 0.074]$). The treatment condition × posttreatment coping interaction ($B = 1.276, p = .753$) added in the moderated mediation model was also not significantly associated with posttreatment symptom severity. The direct effects of treatment condition on outcome at high levels of posttreatment secondary control coping ($B = 0.027, \text{CIs} [-0.364, 0.470]$) and at low levels of posttreatment secondary control coping ($B = -0.069, \text{CIs} [-0.550, 0.336]$) were also not significantly different from zero or from each other ($B = 0.097, \text{CIs} [-0.492, 0.779]$). These findings indicate that parent-reported secondary control coping is predictor but not a mediator of treatment outcome.
**Alliance with youth-reported outcome.** In the simple mediation model, the indirect effect of treatment condition on outcome via alliance was not significant ($B = 0.010$, CIs [-0.031, 0.141]). Treatment condition had no significant effect on late alliance ($B = -0.056$, $p = .724$), which was not significantly associated with posttreatment symptoms ($B = -0.164$, $p = .215$), controlling for early alliance, pretreatment symptoms, and primary diagnosis. Neither the direct effect ($B = 0.252$, CIs [-0.063, 0.586]) nor total effect of treatment ($B = 0.262$, $p = .120$) were significant. In the moderated mediation model, the indirect effects for MCBT ($B = 0.020$, CIs [-0.071, 0.210]) and UC ($B = 0.003$, CIs [-0.028, 0.095]) were not significantly different from zero or from each other ($B = 0.017$, CIs [-0.070, 0.192]). The treatment condition × late alliance interaction ($B = -0.340$, $p = .123$) added in the moderated mediation model was also not significantly associated with posttreatment symptom severity. The direct effect of treatment condition on outcome was not significant when alliance was strong ($B = 0.001$, CIs [-0.468, 0.461]), but was significant when alliance was weak ($B = 0.507$, CIs [0.007, 1.026]), with MCBT resulting in higher symptomatology than UC through mechanisms other than alliance in the context of weak alliance. However, the two direct effects were not significantly different from each other ($B = -0.506$, CIs [-1.211, 0.220]). These findings indicate that alliance is neither a predictor nor mediator of youth-reported outcome.

**Alliance with parent-reported outcome.** In the simple mediation model, the indirect effect of treatment condition on outcome via alliance was not significant ($B = 0.010$, CIs [-0.038, 0.109]). Treatment condition had no significant effect on late alliance ($B = -0.058$, $p = .191$), which was not significantly associated with posttreatment symptom ($B = -0.171$, $p = .191$), controlling for early alliance, pretreatment symptoms, and primary diagnosis. Neither the direct effect ($B = 0.087$, CIs [-0.206, 0.395]) nor total effect of treatment ($B = 0.097$, $p = .539$) were
significant. In the moderated mediation model, the indirect effects for MCBT ($B = 0.012$, CIs $[-0.047, 0.138]$) and UC ($B = 0.008$, CIs $[-0.037, 0.125]$) were not significantly different from zero or from each other ($B = 0.004$, CIs $[-0.047, 0.121]$). The treatment condition × late alliance interaction ($B = -0.062$, $p = .787$) added in the moderated mediation model was also not significantly associated with posttreatment symptom severity. The direct effects of treatment condition on outcome when alliance was strong ($B = 0.041$, CIs $[-0.360, 0.492]$) and when alliance was weak ($B = 0.133$, CIs $[-0.330, 0.616]$) were also not significantly different from zero or from each other ($B = -0.093$, CIs $[-0.716, 0.613]$). These findings indicate that alliance is neither a predictor nor mediator of parent-reported outcome.

**Mediation Models with Therapist Strategies as Predictor**

Youth-reported secondary control coping was the only significant treatment mediator. This section describes analyses conducted to better understand whether cognitive or behavioral strategies may be driving the improvement in outcome via youth-reported secondary control coping regardless of the treatment condition in which the therapists applied those strategies. Two simple mediation models were fitted with cognitive strategies or with behavioral strategies as the predictor, posttreatment youth-reported secondary control coping as the mediator variable, and posttreatment youth-reported symptoms as the outcome, covarying pretreatment coping and symptoms, primary diagnosis, and treatment condition. Table 2.3 displays all point estimates, standard errors, and $p$-values of the models.
Table 2.3. **Regression Models Fitted to Assess Youth-Reported Secondary Control Coping as a Mediator of the Effect of Therapist Cognitive and Behavioral Strategies on Posttreatment Youth-Reported Internalizing Symptoms, Covarying Pretreatment Youth-Reported Symptoms and Secondary Control Coping and Primary Diagnosis, With Parameters Combined Across 50 Imputed Datasets (N = 91)**

<table>
<thead>
<tr>
<th>Predictor/Covariate</th>
<th>Regression Model Predicting Mediator Variable: Posttreatment Youth-Reported Secondary Control Coping</th>
<th>Regression Model Predicting Outcome: Posttreatment Youth-Reported Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.150, SE = 0.036, p &lt; .001</td>
<td>0.687, SE = 0.492, p = 0.164</td>
</tr>
<tr>
<td>Pretreatment Youth-Reported Symptoms</td>
<td>-0.023, SE = 0.008, p = 0.006</td>
<td>0.514, SE = 0.106, p &lt; .001</td>
</tr>
<tr>
<td>Pretreatment Youth-Reported Sec. Control Coping</td>
<td>0.409, SE = 0.162, p = 0.012</td>
<td>0.814, SE = 2.138, p = 0.704</td>
</tr>
<tr>
<td>Primary Diagnosis (0 = depression, anxiety = 1)</td>
<td>0.000, SE = 0.011, p = 0.992</td>
<td>-0.072, SE = 0.141, p = 0.610</td>
</tr>
<tr>
<td>Therapist Cognitive Strategies</td>
<td>-0.043(a), SE = 0.072, p = 0.552</td>
<td>-0.283(c'), SE = 1.075, p = 0.793</td>
</tr>
<tr>
<td>Posttreatment Youth-Reported Sec. Control Coping</td>
<td>-</td>
<td>-6.007(b), SE = 1.679, p &lt; .001</td>
</tr>
</tbody>
</table>

**Predictor: Therapist Behavioral Strategies**

<table>
<thead>
<tr>
<th>Predictor/Covariate</th>
<th>Regression Model Predicting Mediator Variable: Posttreatment Youth-Reported Secondary Control Coping</th>
<th>Regression Model Predicting Outcome: Posttreatment Youth-Reported Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.150, SE = 0.036, p &lt; .001</td>
<td>0.641, SE = 0.492, p = 0.193</td>
</tr>
<tr>
<td>Pretreatment Youth-Reported Symptoms</td>
<td>-0.023, SE = 0.008, p = 0.007</td>
<td>0.522, SE = 0.105, p &lt; .001</td>
</tr>
<tr>
<td>Pretreatment Youth-Reported Sec. Control Coping</td>
<td>0.412, SE = 0.163, p = 0.012</td>
<td>0.962, SE = 2.156, p = 0.656</td>
</tr>
<tr>
<td>Primary Diagnosis (0 = depression, anxiety = 1)</td>
<td>-0.001, SE = 0.011, p = 0.949</td>
<td>-0.066, SE = 0.142, p = 0.645</td>
</tr>
<tr>
<td>Therapist Behavioral Strategies</td>
<td>-0.024(a), SE = 0.108, p = 0.827</td>
<td>0.829(c'), SE = 1.615, p = 0.608</td>
</tr>
<tr>
<td>Posttreatment Youth-Reported Sec. Control Coping</td>
<td>-</td>
<td>-5.962(b), SE = 1.648, p &lt; .001</td>
</tr>
</tbody>
</table>

**Note.** \(B\) = unstandardized regression coefficient, \(SE\) = standard error.

\(a\) Denotes regression coefficient for path \(a\) of mediation model. \(b\) Denotes regression coefficient for path \(b\) of mediation model. \(c\) Denotes regression coefficient for path \(c'\) of mediation model.
**Therapist cognitive strategies.** The indirect effect of cognitive strategies on outcome via youth-reported secondary control coping was not significant ($B = 0.261$, CI [-0.438, 1.234]). Use of cognitive strategies had no significant effect on posttreatment coping ($B = -0.043$, $p = .552$), but more frequent coping was significantly associated with a lower level of symptoms at posttreatment ($B = -6.007$, $p < .001$), covarying pretreatment coping and symptoms as well as primary diagnosis. Neither the direct effect ($B = -0.283$, CI [-2.114, 2.085]) nor total effect of cognitive strategies on outcome ($B = -0.022$, $p = .985$) were significant. These findings indicate therapist cognitive strategies did not drive improvement in youth-reported secondary control coping nor symptom reduction.

**Therapist behavioral strategies.** The indirect effect of behavioral strategies on outcome via youth-reported secondary control coping was not significant ($B = 0.128$, CI [-1.200, 1.321]) as well. Use of behavioral strategies had no significant effect on posttreatment coping ($B = -0.0237$, $p = .827$), but more frequent coping was significantly associated with a lower level of symptoms at posttreatment ($B = -5.962$, $p < .001$), covarying pretreatment coping and symptoms as well as primary diagnosis. Neither the direct effect ($B = 0.829$, CI [-2.702, 3.788]) nor total effect of behavioral strategies on outcome ($B = 0.957$, $p = .566$) were significant. These findings indicate that therapist behavioral strategies did not drive improvement in youth-reported secondary control coping nor symptom reduction.

Given that youth-reported secondary control coping did not mediate the effects of therapist cognitive or behavioral strategies on symptom outcomes, further models to examine whether the mediation effects were moderated by treatment condition, primary diagnosis, or both were not fitted.
Discussion

To understand the mechanisms through which MCBT and UC work under everyday clinical conditions, we examined youth primary control and secondary control coping and therapeutic alliance as mediators of outcome in an effectiveness trial for youths with internalizing symptoms. The only significant mediator was youth-reported secondary control coping; parent-reported primary control and secondary control coping were predictors; and youth-reported primary control coping and youth–therapist alliance were neither mediators nor predictors.

Secondary Control Coping

Although the emergence of secondary control coping as a treatment mediator might have been anticipated, the direction of mediation effects is somewhat surprising. More frequent use of youth- and parent-reported secondary control coping was associated with lower levels of internalizing symptoms, consistent with previous research (Compas, Connor-Smith, Saltzman, Thomsen, & Wadsworth, 2001). However, MCBT led to less frequent secondary control coping than UC. Although this effect was only marginally significant for youth-reported secondary control coping, it resulted in a significant mediation effect. This effect was nonsignificant for parent-report secondary control coping, possibly because the internal nature of most secondary control strategies means that they could only be accessed when verbalized by youths in the parents’ presence. Nevertheless, the same pattern was found where youths who received MCBT rated by their parents as using secondary control coping strategies less frequently than those who received UC. These findings contradict those reported by Compas et al. (2010), in which a group family CBT prevention program increased youths’ secondary control coping, which mediated reductions in depression and internalizing symptoms, compared to a reading materials control.
The present study and that by Compas and colleagues (2010) differ in several respects, but one difference may be key to explaining the discrepant findings. Compas and colleagues used a passive control, precluding any conclusions about whether secondary control coping was a mediator specific to that particular CBT program, specific to CBT more generally, or common to therapies of various orientations. On the other hand, the present study used an active intervention comprising therapist strategies of various orientations as a comparison condition to MCBT, and both increased secondary control coping, which predicted less severe symptoms from pre- to posttreatment. Moreover, therapist use of cognitive and behavioral strategies, regardless of treatment condition, did not increase youth-reported secondary control coping (in fact, they led to a nonsignificant decrease in secondary control coping). These findings are consistent with the hypothesis that secondary control coping may be a common factor driving symptom reduction. However, this common factor hypothesis does not satisfactorily explain why UC led to greater use of secondary control coping than MCBT. Unfortunately, the present study is unable to answer this question. However, it has provided evidence for secondary control coping as a promising candidate change mechanism in therapy for youth internalizing problems worthy of further study.

**Primary Control Coping**

Similar to secondary control coping, both MCBT and UC led to more frequent use of primary control coping strategies, which was associated with better outcomes from pre- to posttreatment. The pre–post change and association were significant only for parent report possibly because the external nature of most primary control coping strategies allowed parents to observe their children’s behavior more easily (and objectively) than the children can observe their own behavior. Nevertheless, the direction of the pre–post change and association was the
same for both informants. Unlike secondary control coping, MCBT and UC did not lead to
differential usage of primary control coping strategies. These findings line up with that of RCTs
that compared CBT to another active intervention and found that both treatment conditions
improved coping (albeit not primary control coping) to an equal extent (e.g., Beidel, et al., 2000).
These findings support the hypothesis that primary control coping may be a common factor of
MCBT and UC for youth internalizing problems.

Therapeutic Alliance

Alliance diverges from both primary and secondary control coping in that it did not
improve from pre- to posttreatment either generally or differentially by treatment condition,
neither did change in alliance predict posttreatment outcomes. These findings do not align with
those of Chiu et al. (2009), who found that improvement in alliance from early to late sessions
predicted superior posttreatment outcomes among anxious youths who received child-focused or
family-focused CBT. Although Chiu and colleagues’ methodology was the most similar to the
present study among other studies of youth-therapist alliance and outcome, there are key
differences. For example, the present study had much larger variation in the number of treatment
sessions and the specific sessions sampled, thus any potential improvement in alliance, and any
association between improvement in alliance and outcome, might have been diluted in the
present study. Interestingly, the direct effect of treatment condition on youth-reported outcome
was significant only when alliance was weak, with MCBT resulting in worse outcomes than UC
through mechanisms other than alliance. A possible explanation is that exposure to feared
stimuli, a core component of CBT for anxiety, increases anxiety initially but leads to habituation
over time within and across episodes of exposure. Poor alliance may have resulted in ineffective
exposure practice in which the youth experienced increased anxiety but did not habituate. Further
research that measures the effectiveness of exposure would be needed to test this hypothesis empirically.

**Therapist Cognitive and Behavioral Strategies**

A notable finding (or lack thereof) is that therapists’ use of cognitive and behavioral strategies neither impacted youth-reported secondary control coping nor internalizing symptoms. This finding is consistent with the common factors hypothesis, but contrasts with RCT research on adult samples (e.g., DeRubeis & Feeley, 1990). Several other explanations are also possible. CBT strategies delivered by the MCBT therapists who may just be learning to use them, or delivered by UC therapists who may be using them in an unstructured manner, may simply not be as potent compared to those delivered by study therapists who are experts in CBT. Another possibility is that therapists are adjusting the amount of CBT strategies they are using according to youth response, so that the relationship between therapist strategies and youth coping or outcome is bidirectional, and associations are muddied. This possibility is consistent with the substantial variation observed in the use of cognitive and behavioral strategies even among MCBT therapists.

**Limitations and Strengths**

An important limitation of the present study is that the mediator and outcome variables were measured concurrently at posttreatment. Without establishing that change in secondary control coping temporally preceded improvement in outcome, it is not possible to conclude that coping is a change mechanism that drove symptom reduction (see Kazdin & Nock, 2003). Alternatives might be symptom reduction driving improvement in coping, or another mechanism driving improvement in both coping and outcome.

In addition, the sample size was small, the proportion of missing data was relatively large.
on the coping and therapist strategy variables, and missingness on some outcome measures may have been MNAR. Handling missing item-level data by prorating (i.e., mean substitution) might have introduced bias towards the mean and could impact the degree of association between variables, and it is unclear how effective multiple imputation at the measure level and compositing of outcome measures were at minimize possible bias in results. Furthermore, bias may have been minimized at the expense of power to detect significant effects, due to the uncertainty introduced by imputing data. The silver lining is that significant effects found over and above the uncertainty are more likely to be reliable effects.

Another issue concerns the appropriateness of computing primary control and secondary control coping scores as proportions of the summed scores of the five dimensions of responses to stress. This approach was taken to adjust for individual differences in response bias, but in retrospect, may have the unintended consequence of magnifying relative differences between each youth’s usage of primary control and secondary control coping responses. For example, a youth who uses both primary control and secondary control coping responses very frequently may receive a moderate proportion score on each coping dimension, whereas another youth who uses primary control coping responses very infrequently, and secondary control coping responses somewhat infrequently may receive proportion scores that are lower in primary control coping and higher in secondary control coping compared to the first youth. More analyses will be needed to explore the relationship between absolute and proportion scores, and whether the results differ depending on how coping scores are computed.

A strength of the present study is that it uses a well-validated measure that distinguishes between two dimensions of coping. In addition, youth- and parent-report versions were administered in the present study. There is good reason to include multiple informants of youths’
coping and symptomatology because inter-informant discrepancies have often been documented in research on youth mental health (De Los Reyes & Kazdin, 2005). Finally, analyses in this study illustrate the application of various quantitative methods that are less frequently used in psychotherapy research, including bootstrapping in combination with multiple imputation to test both simple and moderated mediation models.

**Future Directions**

The results of the present study suggest several directions for future research. First, secondary control coping, particularly as self-reported by youths, warrants further study as a candidate change mechanism. The issue of temporal precedence can be addressed by measuring coping and symptoms multiple times during treatment as well as posttreatment. Intensive and repeated measurements were previously considered too burdensome and expensive, but are now much more feasible with the advent of mobile technologies, as evidenced by the burgeoning field of ecological momentary assessment (EMA; Shiffman, Stone, & Hufford, 2008). EMA is just starting to be utilized in RCTs of youth psychotherapy (e.g., Silk et al., 2016) and is likely to produce critical information about temporal relationships among outcomes, candidate change mechanisms, and therapy processes. An emerging methodology that may be especially useful in the absence of multiple repeated measurements is causal mediation analysis (Imai, Keele, & Tingley, 2010; Valeri & VanderWeele, 2013). This method has the advantage of identifying mediators that are more likely to cause the outcome by quantifying the robustness to confounding of the mediator–outcome relationship. Evidence that youth-reported secondary control coping is robust to mediator–outcome confounding would provide stronger empirical support for this mediator as a promising candidate change mechanism.

Second, it remains unclear what aspects of UC make it more effective than MCBT at
enhancing secondary control coping skills among youths. UC therapists often engaged the youths in reflecting on or understanding their experiences and feelings, which could have fostered greater acceptance of stressors that cannot be easily changed. On the other hand, acceptance is one aspect of secondary control coping that is not typically emphasized in CBT for youth depression and anxiety, including in the protocols used in the MCBT condition of the present study. Instead, most CBT protocols focus on helping youths change their thinking and behavior, which may be incompatible with some aspects of secondary control coping. RCTs of acceptance and mindfulness-based therapies are beginning to be published, and initial studies suggest that such therapies may be comparable to CBT in efficacy (e.g., Hancock et al., 2016). Alternatively, UC therapists may be freer to tailor therapy to individual youths, helping them to strengthen existing secondary control coping skills, to build new ones, and to select those they find most effective, whereas MCBT therapists may have felt limited in their ability to individualize therapy, especially if they were new to CBT. This idea is consistent with two lines of our previous work indicating that youths with higher levels of depression do perceive certain coping strategies as effective but are less likely to use them habitually (Ng, Eckshtain, & Weisz, 2015); and that a modular EBP protocol that facilitates personalization of therapy outperformed standard EBPs for youths with depression, anxiety, conduct problems, or some combination of the above (Weisz et al., 2012). Systematic study of what goes on in UC sessions will shed light on how UC may be helping youths build up secondary control coping skills. Indeed, “practice-based evidence” (Garland, Hurlburt, & Hawley, 2006), including monitoring the contents of usual care, has gained momentum in the field of psychotherapy research. It is possible that information about the contents of usual care could be used to improve CBT.

Third, what are the candidate mechanisms through which MCBT works to improve
symptom outcomes in everyday clinical conditions? The fact that Weisz et al. (2009) and Southam-Gerow et al. (2010) found no between–group differences in symptom outcomes, suggests that MCBT may act on some mechanism other than primary or secondary control coping or therapeutic alliance to bring about symptom reduction. The search for mediators in RCTs of CBT for youth anxiety and depression has proved to be rather elusive, with negative cognitions and anxious self-statements surfacing among the few significant mediators (for a review, see Weisz, Ng, Rutt, Lau, & Masland, 2013). One might expect negative cognitions and anxious self-statements to be closely and negatively associated with secondary control coping, but this assumption would need to be examined empirically.

Although CBT is generally efficacious for treating youth internalizing disorders, the effect sizes generated by these treatments tend to be quite modest, particularly when compared to UC (e.g., Weisz et al., 2013). By examining mediators of MCBT and UC in everyday clinical conditions, psychotherapies may be refined to maximize treatment gains for youths with internalizing disorders.
References


Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology


Dissertation Study 3:
Clarifying the Direction of Change Between Personalized “Top Problems” and Symptom Outcome Trajectories in Modular vs. Standard Evidence-Based Psychotherapies for Internalizing and Externalizing Youths

Mei Yi Ng
Harvard University
Abstract

**Objective:** Many evidence-based psychotherapies (EBPs) for youths have demonstrated impressive efficacy in research trials, only to falter in clinically representative conditions. Modular EBPs were designed to facilitate therapists’ tailoring of EBPs to meet the diverse needs of treatment-seeking clients. The present study investigated whether improvement in parent-identified “top problems”—a measure of youths’ achievement of personalized treatment goals—may underlie the superiority of modular EBPs over standard EBPs in a recent effectiveness trial.

**Method:** Bivariate latent change score models were fitted to examine reciprocal relationships between top problems and symptom outcome among 120 youths with elevated internalizing and externalizing symptoms over 16 weeks of treatment with either modular or standard EBPs. Analyses examined whether weekly top problem severity predicted subsequent symptom improvement (i.e., whether top problems is a leading indicator), whether the reverse pattern held (i.e., whether symptom severity is a leading indicator), and whether the leading indicator(s) mediated the effects of treatment condition on the other variable. **Results:** Model fit was only marginally acceptable, thus findings are tentative. Symptom severity significantly predicted subsequent changes in top problems, whereas the reverse pattern was not supported. Symptom change did not significantly mediate treatment effects on top problem severity. **Conclusion:** Preliminary findings did not support top problems as a change process by which modular EBPs outperformed standard EBPs. Instead, symptom reduction appeared to drive top problem improvement across modular and standard EBPs. Emerging methods for intensive longitudinal data may be needed to accurately model the dynamics between change processes and treatment outcomes.
Keywords: children and adolescents, modular therapy, evidence-based psychotherapy, personalized intervention, Top Problems
Study 3: Clarifying the Direction of Change Between Personalized “Top Problems” and Symptom Outcome Trajectories in Modular vs. Standard Evidence-Based Psychotherapies for Internalizing and Externalizing Youths

Evidence-based psychotherapies (EBPs) earn their designation by demonstrating their efficacy in randomized controlled trials (RCTs). However, their effectiveness typically suffers when implemented under conditions representative of real-world clinical practice, and when compared to usual clinical care. This drop in therapeutic benefit when transporting an EBP from research to clinical settings has been termed the “implementation cliff” (Weisz, Ng, & Bearman, 2014). The implementation cliff may stem from the myriad ways in which RCTs and everyday clinical practice differ (see Weisz, Krumholz, Santucci, Thomassin, & Ng, 2015; Weisz, Ugueto, Cheron, & Herren, 2013). For example, RCTs are usually focused on treatment of a single disorder or homogeneous symptom cluster, and they typically involve treatment of recruited individuals who only, or primarily, have that disorder or those symptoms. EBPs are usually delivered via single-disorder manuals that standardize the content and structure of therapy to improve the internal validity of the RCT. In contrast, practicing clinicians often need to treat a variety of diagnoses and problems across and within individuals. Indeed, treatment-seeking youth populations tended to display high heterogeneity of problems across individuals, high comorbidity within individuals, and flux in problems over time, including crises that arise, and treatment-interfering behaviors or circumstances. Thus many clinicians feel that EBPs developed and tested in RCTs are not very generalizable to their clients (Stewart, Stirman, & Chambless, 2012). Research suggests that clinicians are not opposed to evidence-based practices so much as to the rigid, “cookbook-like” structure of manuals (Borntrager, Chorpita, Higa-McMillan, & Weisz, 2009; Thomas, Zimmer-Gembeck, & Chaffin, 2014). In fact, clinicians have indicated...
greater willingness to use treatments that incorporate flexibility to address the severe and diverse problems they see in their clients (Nelson, Steele, & Mize, 2006).

**Modular vs. Standard Evidence-Based Psychotherapies**

To address this mismatch between the design of EBPs and the needs of treatment-seeking clients, Chorpita and Weisz (2005) developed *Modular Approach to Therapy for Children with Anxiety, Depression, or Conduct Problems* (MATCH; later published as Chorpita & Weisz, 2009). MATCH contains content (i.e., therapy practices) patterned after that found in standard EBPs for youth anxiety, depression, and disruptive behavior. The content in standard EBPs is organized in a linear design, with practices sequenced in a consistent order, session-by-session, and with separate treatment manuals for each diagnostic category or homogenous cluster. The client’s primary diagnosis determines which treatment manual will be used, and the practices are intended to be delivered to all clients following the fixed session-by-session structure. In MATCH, EBP content is re-organized into a modular design such that practices for multiple diagnostic categories are organized into self-contained modules that can be used multiple times or not at all, and can be combined with one another as needed; the client’s primary diagnosis or problem is used to select a clinical-decision-making flowchart that takes into account the client’s response to treatment, presence of comorbid problems, and emergence of treatment-interfering behaviors (see Weisz & Chorpita, 2011). Whereas two clients who completed a course of the same standard EBPs would receive the same practices in the same, fixed sequence over roughly the same number of sessions, two clients who completed MATCH might receive several of the same core modules, possibly over a different number of sessions, and could receive different additional modules, no additional modules, or a different sequence of the same modules.
In a randomized trial, Weisz and colleagues (2012) compared the effectiveness of MATCH to standard EBPs (STEBP) and usual care (UC) among youths with elevated internalizing symptoms (i.e., anxiety, depression), externalizing symptoms (e.g., conduct problems, disruptive behavior), or some combination of thereof. This RCT was an effectiveness study rather than an efficacy study because all three treatments were delivered by practicing therapists in community- or school-based outpatient mental health service settings and with treatment provided to youths who had been referred for treatment through usual community channels (rather than being recruited, e.g., through advertisements). Therapists were randomized to treatment conditions, and those assigned to MATCH and STEBP received training in their assigned treatments and supervision from study staff, who used a web-based tool to monitor client progress and therapy content covered. Weisz and colleagues (2012) found that MATCH outperformed STEBP and UC on change trajectories of symptom severity and posttreatment diagnostic outcomes, whereas STEBP and UC did not differ on outcomes. Treatment gains of MATCH relative to UC were maintained at two-year follow-up (Chorpita et al., 2013).

The randomized effectiveness trial demonstrated that the modular MATCH protocol conferred advantages over STEBP, yielding greater therapeutic benefit. But it remains to be tested empirically whether the superior effects of MATCH were due to its utility in guiding therapists to select specific EBP practices and deliver them in a sequence that is customized to meet the needs of individual youths and their parents. This is a challenging hypothesis to test for various reasons. For one, it is not apparent how one would assess the constellation of unique needs of individual youths in a way that can be compared to others in the sample, and then operationalize the effective customization of EBP practices by therapists to meet those needs. For another, even if one were to zoom in on a specific need such as the presence of comorbidity,
and to assess the therapists’ use of EBP content to target comorbid disorders, it is doubtful that frequency or extensiveness of the usage of any particular practice would have meaningful associations with outcome. Due to the issue of responsiveness—adjusting the usage of treatment practices in response to clients’ needs—therapists might intensify or dampen a practice to achieve optimal outcomes for each client (see Doss, 2004). After all, modular therapies were designed to support therapists to respond to individual clients’ progress, or lack thereof, skillfully. Unfortunately, therapist skill is also difficult to measure, resulting in a dearth of psychometrically sound measures in the youth psychotherapy literature (Weisz, Ng, Rutt, Lau, & Masland, 2013). More progress has been made in measuring therapist competence in adult psychotherapy, but existing measures have not shown a reliable association with treatment outcome (Webb, DeRubeis, & Barber, 2010).

**Assessing Personalized Treatment Goals**

Clearly, assessing therapists’ skill at tailoring treatment to individual clients is a complex endeavor; but one can begin by examining its more accessible proxy—clients’ achievement of personalized treatment goals. Measures of personalized goal achievement have long been available, such as Goal Attainment Scaling (GAS; Kiresuk & Sherman, 1968). However, RCT researchers have often eschewed such idiographic measures in favor of nomothetic assessments, paralleling their emphasis on standardized over individualized treatment delivery, in service of producing replicable results (Persons, 1991). A recent meta-analysis (Lindhiem, Bennett, Orimoto, & Kolko, 2016) found only 12 psychotherapy trials that assessed personalized treatment goals in addition to standardized symptom measures. In this sample of studies, effect sizes computed from personalized treatment goals were significantly larger than those computed from standardized symptom measures, generating empirical support for earlier suggestions that
idiographic measures of change mechanisms or outcome may be more sensitive measures of clinical change than standardized measures (Doss, 2004).

**Study Aims**

The present study aim was to use data from the Weisz et al. (2012) effectiveness trial to the investigate achievement of personalized treatment goals as a candidate change process underlying the success of modular over standard EBPs in improving overall symptom outcome. Personalized goal achievement was assessed by the Top Problems Assessment (TPA; Weisz et al, 2011), through which youths and parents separately identify the problems they deem most important to address at the beginning of treatment, then rate the severity of each “top problem” weekly for the duration of treatment. The outcome measure was the Brief Problem Checklist (BPC; Chorpita et al., 2010), a measure of internalizing and externalizing symptoms administered weekly to youths and parents alongside the TPA. In the idiographic measure of top problems, each participant rates a unique set of self-generated items, whereas in the nomothetic symptom measure, each informant rates the same set of researcher-generated items. The primary outcome paper for the Weisz et al. (2012) trial reported that both top problems and symptom outcomes rated by youths and parents declined more steeply for MATCH than STEBP.

Bivariate latent change score models (LCS; McArdle, 2009) were fitted to examine whether weekly top problem severity predicted subsequent symptom change and whether weekly symptoms predicted subsequent change in top problem severity among youths receiving up to 16 weeks of modular or standard EBPs. If either or both top problems and symptoms prospectively predicted change in the other variable, that is, if it was a leading indicator, then we also tested whether the leading indicator mediated the effects of treatment condition on subsequent change in the other variable. We considered both youth- and parent-reported top problems and symptom
outcomes for analyses but ultimately excluded the youth-reported measures from analyses due to difficulty in modeling the change trajectories of those measures with adequate fit (see Data Analyses section for more details).

The hypothesis that modular EBPs produce greater treatment gains than standard EBPs by better supporting therapists to help youths achieve personalized treatment goals would be supported by the following findings: (a) top problems was a leading indicator of symptom change, (b) MATCH led to greater reduction in top problem severity, and (c) top problems mediated the effects of MATCH on symptom change relative to STEBP. Alternatively, if top problems was a leading indicator of symptom change with no difference in top problem severity reduction between treatment condition, and no significant mediation effect, then top problems could be a change process common to modular and standard EBPs. Another alternative is if symptoms were a leading indicator of change in top problem severity. This finding would suggest that treatments reduce symptoms through other change processes and mechanisms, and that overall symptom reduction might either enable youths to achieve specific treatment goals identified by their parents, or make parents aware that the treatment goals they nominated for their child were being met.

Method

Participants

Participants were 120 youths seeking treatment at ten community outpatient programs, as well as their caregivers (to whom we refer as “parents”), referred through normal community channels. Therapists were assigned to treatment conditions via cluster randomization procedures (Campbell, Elbourne, & Altman, 2004; Dohner & Klar, 2000), and youths received the treatment condition assigned to their therapists. Of the 174 youths in the primary outcome paper (Weisz, et
al., 2012), the present study included 61 of the 62 youths who received MATCH and all 59 youths who received STEBP. The 53 youths who received UC were excluded, as was one youth who received MATCH but did not have a weekly assessment following any treatment session. Youths were aged 7 – 13 years; 31.7% were female; 49.2% of participants were Caucasian American, 27.5% were of mixed ethnicity, 8.3% were African American, 7.5% were Latina/Latino, 5.8% were Asian/Pacific Islander, and 1.7% chose not report their ethnicity. Youths were included if they had a primary problem (i.e., diagnosis or elevated symptoms) of anxiety, depression, or disruptive behavior. Diagnoses were generated using the Children’s Interview for Psychiatric Syndromes, child (ChIPS) and parent (P-ChIPS) versions (Weller, Weller, Rooney, & Fristad, 1999a, 1999b). Elevated symptoms (T-score ≥ 65, Borderline or Clinical range) were assessed with the Child Behavior Checklist (CBCL) and Youth-Self Report (YSR; Achenbach & Rescorla, 2001). Comorbid problem types included the primary problem disorder categories as well as eating disorders, elimination disorder, and bipolar disorder. Youths with psychotic symptoms, pervasive developmental disorder, mental retardation, or primary problems of inattention/hyperactivity or bipolar disorder were excluded. Youths did not differ significantly on the number of disorders across conditions.

Treatment Conditions

**Standard evidence-based psychotherapy.** STEBP therapists were trained to use three manualized EBPs which prescribe a fixed, linear sequence of treatment components. In *Coping Cat* (Kendall, Kane, Howard, & Siqueland, 1990), an individual CBT protocol for youth anxiety, youths engage in identification and reframing of anxious thoughts, identification and alleviation of physical sensations through relaxation, problem solving, self-monitoring and reinforcement of approach behavior, and graduated exposure to fear-inducing stimuli over 16 – 20 sessions. In
Primary and Secondary Control Enhancement Training (PASCET; Weisz et al., 2005), an individual CBT protocol for youth depression, youths engage in problem solving, pleasant activity scheduling, relaxation, positive self-presentation, talent building, identification and reframing of negative thoughts, positive thinking, thought stopping, and social support-seeking over 10 – 15 sessions. In Defiant Children (Barkley, 1997), a behavioral parent training protocol for youth disruptive behavior, parents engage in regularly scheduled one-on-one time for relationship-building with their child, selective attention to and reinforcement of nondisruptive or compliant behavior, time out for disruptive or noncompliant behavior, and consistent delivery of consequences across settings over 10 “steps” (roughly corresponding to sessions). All three protocols emphasize therapist modeling of skills, as well as client practice of skills via in-session role plays and activities, and between-session homework assignments.

Modular Approach to Therapy for Children with Anxiety, Depression, or Conduct Problems. MATCH (Chorpita & Weisz, 2005; Weisz & Chorpita, 2011) therapists were trained to use a single protocol comprising modules corresponding to the components of three manuals for youth anxiety, depression, and disruptive behavior. Three decision flowcharts, each designed for one primary problem but containing modules from the EBPs of all three problem areas, guide therapists’ use of modules in a flexible manner. The specific modules delivered, and the sequence in which they are delivered depends on the youths’ response to treatment, presence of comorbid problems, and emergence of treatment-interfering behaviors. Although two youths with the same primary problem would receive the same core modules, they might receive the same modules over a different number of sessions, receive different additional modules or no additional modules, and receive modules in a different sequence.

STEBP and MATCH therapists were trained together for six days; they received the same
training on the content of modules/components but participated in separate group discussions on using the MATCH decision flowcharts to select modules vs. using the STEBP protocols to carry out the components in the prescribed sequence (see Borntrager et al., 2009). They also received weekly supervision from 12 doctoral-level research staff members (see Bearman et al., 2013). Supervision was facilitated by supervisors’ review of each client’s progress and their therapist’s coverage of treatment content via an Internet-based monitoring and feedback tool (Chorpita, Bernstein, Daleiden, & the Research Network for Youth Mental Health, 2008). Adherence coding of session audio- or videotapes showed that MATCH and STEBP therapists delivered their assigned treatments as intended (Weisz et. al, 2012).

**Measures**

Upon assignment of each youth to a therapist, research assistants blinded to treatment condition made weekly phone calls to youths and parents to administer the following assessments. The calls were continued until the youth completed treatment or dropped out of the study.

**Brief Problem Checklist.** The BPC (Chorpita et al., 2010) is a 12-item interview with parallel parent- and youth-report versions that was designed for weekly assessment of youth internalizing and externalizing problems. It was developed via item response theory and factor analysis of data from the CBCL and YSR, which are widely used parent- and youth-report symptom measures with strong psychometric properties (Achenbach & Rescorla, 2001). The validation study indicates that the BPC is highly reliable and valid (Chorpita et al., 2010); notably, the BPC is strongly correlated with the CBCL and YSR, and also more reliable when administered on a weekly basis than the CBCL and YSR administered on a three-monthly basis on some reliability indices. In addition, the BPC predicts change in other measures of youth
symptoms.

**Top Problems Assessment.** The TPA (Weisz et al., 2011) is an interview with parallel parent- and youth-report versions that was designed to identify up to three problems of greatest concern to each informant at pretreatment, and then to track the severity of each problem weekly. As part of the validation study, parent- and youth-generated top problems were coded for their correspondence to CBCL and YSR items (Weisz et al., 2011). Top problems largely corresponded to CBCL and YSR items while adding more specific information, demonstrating the TPA’s clinical relevance and incremental validity. For example, a parent or youth might endorse the item “worries” on the CBCL or YSR, but identify the content and nature of the worries in the TPA (e.g., worries about mother, being away from parents, “doom,” failure, being perfect, bad grades, adoption, things out of the youth’s control, “the future,” “everything”). Other top problems were broader than the symptoms of internalizing and externalizing disorders (e.g., communication, lack of social skills, reading difficulties, emotional instability, having a hard time with parents’ divorce). TPA scores, computed from summing the severity ratings across the three top problems of each informant, were shown to be reliable, valid, and sensitive to change over time during the course of treatment.

**Data Analyses**

**Binning of treatment sessions and weekly assessments.** Because weekly assessment calls were made independently of treatment sessions and depended on the families’ schedules, there was considerable variation in the timing of each session in relation to each assessment across weeks and participants. To structure the data for analyses, dates of documented treatment sessions and weekly assessments were binned into weeks using an approach that maximized the number of weekly bins containing a session that was paired with an assessment in the subsequent
bin. This approach involved creating seven bin variants for every youth, each starting with a different day of the week. The number of weekly bins containing at least one treatment session with a paired assessment in the next bin was summed up per participant, and the bin variant with the highest sum of treatment–assessment pairs was selected for each participant. If multiple bin variants had the highest sum of treatment–assessment pairs, then the bin variant starting on an earlier day of the week was selected by default (e.g., Monday had priority over Tuesday, which had priority over Wednesday). If a bin contained multiple weekly assessments, then any assessment that had both symptom and top problem scores was chosen over any assessment that had a missing score on one of the two measures. If there was more than one assessment with both scores in that bin, then the earlier assessment was selected in order to reduce the time lag between that week’s assessment and the treatment session in the previous bin. This approach ensured that each treatment session was paired with an assessment that came after it. Bins were numbered according to the first documented treatment session such that week 1 corresponded to the week the first treatment session occurred. Throughout the rest of this manuscript, the week number refers to the number of weeks since the first treatment session, which is also the number of weeks of treatment. Thus week 8 is the eighth week since the first treatment session, and is also the eighth week of treatment.

Selection of time points. There was substantial variation in treatment duration, ranging to from 1 to 82 weeks across the 120 participants (MATCH $M=28.3$, $SD=16.8$; STEBP $M=26.2$, $SD=16.3$). The first 16 weeks of treatment was identified to adequately represent a course of standard EBP (see description of typical duration above), the actual treatment received by most study participants (80.3% of MATCH participants and 69.5% of STEBP participants had at least 16 weeks of treatment), as well as the average treatment duration in the largest meta-analysis of
youth psychotherapy RCTs to date (Weisz et al., 2017)—while balancing the need for sufficient data in each week to model change trajectories. Accordingly, assessments conducted in the week of the first treatment session plus those in up to 16 subsequent weeks were selected for analyses. Participants who had treatment sessions through week 16 or beyond had assessment data from weeks 1 through 17 included. Participants who had their last session before week 16 had data from weeks 1 up through the week containing the earliest assessment following their last session included, as long as that session occurred in week 17 or earlier. For example, if a participant had her last session in week 10, missed her assessments in weeks 11 and 12, and had her next assessment in week 13, then her assessments from weeks 1 through 13 were included. Any assessment conducted after the earliest assessment following the last treatment session were excluded from analyses, as were any assessments conducted before week 1.

**Computing potential covariates.** After participants were assigned to therapists, it was up to therapists and families to schedule their first treatment session and termination session, and some families dropped out of treatment while remaining in the study. The weekly assessments were a separate procedure conducted by study staff, thus leading to some discrepancies in the numbered weeks that the first or last treatment session occurred in relation to the first or last weekly assessment. For example, some participants had their first assessment before week 1 (of treatment), others had their first assessment during week 1, and yet others had their first assessment after week 1. Because it is unclear whether there were any systematic factors causing some participants to have their first assessment before vs. after their first treatment session, the number of weeks that the first assessment came after the first treatment was computed as a potential covariate in case this variable had any effects on the key variables of interest. Similarly, the number of weeks between the last treatment session, if it was before week 16, and the
following assessment, if it was within week 17, was computed as a covariate. Participants who had their last treatment session on week 16 or later received a value of 1 on this covariate to represent the latest treatment session paired with a next-week assessment within the included weeks, 1 through 17.

**Preliminary modeling of outcome trajectory.** The Loess curves of symptom outcomes for individual participants and for each treatment condition were examined, and these indicated a nonlinear change trajectory for many participants, with substantial between-individual variation. A variety of longitudinal growth models were fitted separately to parent-reported and youth-reported symptoms from weeks 1 through 17 to identify the overall functional form of the symptom outcomes and to aid in selection of a suitable type of growth model for the present analyses. The following types of models were fitted: latent growth curve (LGC) models with linear, quadratic, cubic, quartic, and quintic slopes, with and without autoregressive paths (i.e., paths from each week’s outcome predicting the next week’s outcome); piecewise LGC models with the trajectories divided into two or three pieces at fixed knot points and at estimated knot point (i.e., empirically determine where to break the trajectory into two pieces), with linear, quadratic, cubic, and quartic slopes estimated for each piece, without and without autoregressive paths; autoregressive models with up to $10^{th}$ order autoregressive effects (i.e., paths from week 1 outcome predicting outcome at each of the next 10 weeks, paths from week 2 outcome predicting outcome at each of the next 10 weeks and so on); latent basis growth models; and LCS models (detailed below). These models are described in texts on longitudinal growth modeling (e.g., Grimm, Ram, & Estabrook, 2017). Most of the models fit the data poorly or resulted in
estimation issues (e.g., nonpositive definite⁷ first-order derivative product matrix, likely due to more parameters being estimated than available sample size, leading to standard errors of parameters that may be inaccurate). Only two models were free of estimation issues and had marginally acceptable fit to the parent-reported symptoms: the LGC model with quadratic slope and autoregressive paths, and the LCS dual change plus dynamic noise model (described below). None of the models fit the youth-reported symptoms adequately, thus this outcome was excluded from analyses.

Latent change score models. A series of LCS models, also known as latent difference score models, were fitted to test the present study’s hypotheses. LCS modeling decomposes an individual’s observed score at a particular time point into his true score at the first time point, accumulated changes in true score that have occurred between each pair of time points up to that particular time point, and a unique score for that individual at that particular time point (Grimm et al., 2017; McArdle, 2009). The LCS framework facilitates evaluation of whether two repeatedly measured variables each predicts later change in the other variable; on the other hand, LGC models test whether the growth curve of one variable is associated with that of another variable (Ferrer & McArdle, 2003). Given that LCS models are more suited to clarifying the direction of change between variables, and the preliminary LCS and LGC models showed comparable (albeit marginally acceptable) fit, analyses proceeded with LCS models.

Univariate latent change score models. First, a set of univariate LCS models were fitted to parent-reported symptoms and their relative and absolute fit were compared. The models differed in the shape of change specified, and included (a) no change, (b) constant change (i.e., constant rate of change, producing a linear slope), (c) proportional change (i.e., predicted

⁷ Positive definiteness is a characteristic of a matrix that is required for SEM and most other multivariate analyses. It involves several features, such as being invertible, having all positive eigenvalues, and having no out-of-bounds correlations and covariances (see Kline, 2011). A matrix that does not have these features is nonpositive definite.
changes are proportional to the level of the prior true score), (d) dual change (comprising constant change and proportional change, which leads to an exponential slope), (e) linear change (i.e., linear rate of change, which is equivalent but not identical to a quadratic slope), (f) linear change plus dual change (see Grimm et al., 2017). Dynamic noise, a residual term added to the change equation such that the latent change score is not perfectly predicted by model parameters, was incorporated into some of the models (McArdle, 2001). Second, the same set of univariate LCS models were fitted to parent-reported top problems. The best fitting univariate models for parent-reported symptoms and for parent-reported top problems were selected for inclusion in the bivariate LCS models. Youth-reported top problems were also fitted to the univariate LCS models but none of the models fit the data adequately; this measure was therefore excluded from further analyses.

**Bivariate latent change score models.** A third set of models combined the two best-fit univariate models for parent-reported symptoms and top problems to examine four possible patterns of reciprocal relations: (a) no coupling (i.e., neither variable predicts subsequent change in the other), (b) unidirectional coupling whereby top problems predicts subsequent symptom change, (c) unidirectional coupling whereby symptoms predict subsequent change in top problem severity, and (d) full (i.e., bidirectional) coupling whereby each variable predicts changes in the other variable (see Grimm et al., 2017). The best fitting model was selected for the incorporation of covariates and treatment condition.

**Models with covariates and treatment condition added.** This fourth set of models added the two computed covariates and treatment condition. Direct and indirect effects of treatment condition were estimated, with the growth factor of the leading indicator(s) modeled as the candidate mediator and the weekly latent change scores of the other variable modeled as the
outcome. Covariates and paths that were nonsignificant, caused estimation issues, or did not improve model fit were pruned from the model. The significance of indirect effects was tested using the delta method (Sobel, 1982). 

Descriptive analyses were conducted with SAS version 9.4. Longitudinal growth modeling was conducted with Mplus 8 (Muthén & Muthén, 2017). Full information maximum likelihood estimation (FIML) was used; FIML has been recommended for handling missing data in longitudinal analyses because available data at some time points can be used to predict missing data at other time points (Graham, 2009). Robust maximum likelihood (MLR) estimation of standard errors was employed because histograms revealed that the distribution of top problems and symptoms departed from normality in some weeks (e.g., skewed right, higher frequency of low scores especially in later weeks). Relative fit of nested models was assessed with Crayen’s (2010) chi-square difference calculator.

**Results**

**Descriptive Statistics**

Table 3.1 displays the raw means, standard deviations, sample sizes, and zero-order correlations for all variables used in the models. Sample sizes for the symptom scores and top problems scores each week ranged from 67 to 101 ($M = 79.76$). Consistent with the findings of the Weisz et al. (2012) primary outcome paper, mean symptom and top problem scores appear to be decreasing from weeks 1 to 17 of treatment. The computed covariates indicated that on average, the first assessment occurred 1.15 weeks before the first treatment session, and the last assessment occurred 1.11 weeks after the last treatment session within weeks 1 to 17 of treatment.

---

8 Attempts to compute 95% bias-corrected bootstrapped confidence intervals, which offer several advantages over the delta method (Preacher & Hayes, 2008) were not successful due to limits on computer memory and space, as well as the large number of paths involved in the indirect effect in the bivariate LCS model specified.
As expected, symptoms scores and top problem scores were positively correlated within and between measures, and within and between weeks, and most correlations were significant ($p < .05$). Symptom scores were moderately to highly correlated across weeks (mean $r = .68$, range $= .34 – .91$), with correlations getting weaker the further apart the weeks. A similar pattern of moderate to high correlations was found for top problem scores across weeks (mean $r = .67$, range $= .27 – .87$). Across all weeks, correlations between symptom and top problem scores were low to moderate (mean $r = .37$, range $= .04 – .61$), with lower correlations once again occurring the further apart the weeks. The subset of within-week correlations between symptom and top problem scores were moderate (mean $r = .54$, range $= .48 – .61$).

Point biserial correlations between treatment condition (MATCH = 1, STEBP = 0) and symptoms were not significant in any week, neither were those between treatment condition and top problems (all $p > .05$). However, the sign of the correlations indicates that symptoms and top problem severity were higher for MATCH than STEBP for the first 8 to 11 weeks, and then subsequently became lower for MATCH, compared to STEBP. This pattern is consistent with the primary outcome paper finding of steeper declines in symptom and top problem severity for MATCH than for STEBP (Weisz et al., 2012).

The two covariates were not significantly correlated with each other or with treatment condition. Correlations with symptoms and top problems were small and there was a mix of negative and positive correlations. The “first treatment–assessment discrepancy” covariate was not significantly correlated with symptom scores at any week (mean $r = .03$, range $= -.11 – -.12$); it was significantly ($ps < .05$) and negatively correlated with top problems at weeks 4 ($r = -.21$), 9 ($r = -.21$), and 17 ($r = -.26$; mean $r = -.15$, range $= -.26 – -.01$). The “last treatment–assessment discrepancy” covariate was also not significantly correlated with symptom scores at any week.
(mean $r = -.11$, range $= -.20 – .003$); it was significantly ($p < .05$) and negatively correlated with top problems scores at week 4 ($r = -.30$; mean $r = -.06$, range $= -.30 – .09$). Thus participants who had a longer lag of their first weekly assessment following their first documented treatment session, and who had a longer lag between their last treatment session and next assessment, tended to have lower top problems scores at certain weeks.
<table>
<thead>
<tr>
<th></th>
<th>Tx</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>B4</th>
<th>B5</th>
<th>B6</th>
<th>B7</th>
<th>B8</th>
<th>B9</th>
<th>B10</th>
<th>B11</th>
<th>B12</th>
<th>B13</th>
<th>B14</th>
<th>B15</th>
<th>B16</th>
<th>B17</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M</strong></td>
<td>0.51</td>
<td>8.70</td>
<td>9.10</td>
<td>8.32</td>
<td>8.48</td>
<td>9.19</td>
<td>8.16</td>
<td>7.88</td>
<td>8.79</td>
<td>8.25</td>
<td>8.27</td>
<td>7.36</td>
<td>6.94</td>
<td>5.89</td>
<td>7.03</td>
<td>5.75</td>
<td>7.20</td>
<td>6.82</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>0.50</td>
<td>4.75</td>
<td>4.76</td>
<td>4.96</td>
<td>5.02</td>
<td>5.33</td>
<td>5.31</td>
<td>4.68</td>
<td>5.28</td>
<td>5.25</td>
<td>4.73</td>
<td>5.18</td>
<td>5.19</td>
<td>4.61</td>
<td>5.27</td>
<td>4.74</td>
<td>5.14</td>
<td>5.16</td>
</tr>
<tr>
<td><strong>n</strong></td>
<td>120</td>
<td>73</td>
<td>101</td>
<td>94</td>
<td>88</td>
<td>83</td>
<td>91</td>
<td>88</td>
<td>82</td>
<td>84</td>
<td>70</td>
<td>73</td>
<td>72</td>
<td>70</td>
<td>74</td>
<td>75</td>
<td>71</td>
<td>67</td>
</tr>
</tbody>
</table>

Table 3.1 *Raw Sample Means, Standard Deviations, Sample Sizes, and Zero-order Correlations (N=120)*
Table 3.1 (Continued).

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
<th>T8</th>
<th>T9</th>
<th>T10</th>
<th>T11</th>
<th>T12</th>
<th>T13</th>
<th>T14</th>
<th>T15</th>
<th>T16</th>
<th>T17</th>
<th>C1</th>
<th>C17</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>6.42</td>
<td>6.60</td>
<td>6.28</td>
<td>6.20</td>
<td>6.20</td>
<td>6.14</td>
<td>5.75</td>
<td>5.80</td>
<td>5.49</td>
<td>5.55</td>
<td>5.12</td>
<td>4.97</td>
<td>4.73</td>
<td>5.14</td>
<td>4.19</td>
<td>5.06</td>
<td>4.57</td>
<td>-1.15</td>
<td>1.11</td>
</tr>
<tr>
<td>SD</td>
<td>2.25</td>
<td>2.30</td>
<td>2.35</td>
<td>2.42</td>
<td>2.48</td>
<td>2.52</td>
<td>2.40</td>
<td>2.40</td>
<td>2.65</td>
<td>2.65</td>
<td>2.73</td>
<td>2.51</td>
<td>2.49</td>
<td>2.53</td>
<td>2.70</td>
<td>2.58</td>
<td>2.68</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>73</td>
<td>101</td>
<td>94</td>
<td>88</td>
<td>84</td>
<td>93</td>
<td>89</td>
<td>87</td>
<td>90</td>
<td>76</td>
<td>78</td>
<td>77</td>
<td>72</td>
<td>74</td>
<td>75</td>
<td>71</td>
<td>67</td>
<td>120</td>
<td>120</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Tx</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>B4</th>
<th>B5</th>
<th>B6</th>
<th>B7</th>
<th>B8</th>
<th>B9</th>
<th>B10</th>
<th>B11</th>
<th>B12</th>
<th>B13</th>
<th>B14</th>
<th>B15</th>
<th>B16</th>
<th>B17</th>
</tr>
</thead>
<tbody>
<tr>
<td>T8</td>
<td>0.03</td>
<td>0.25</td>
<td>0.34</td>
<td>0.61</td>
<td>0.32</td>
<td>0.43</td>
<td>0.32</td>
<td>0.45</td>
<td>0.51</td>
<td>0.32</td>
<td>0.31</td>
<td>0.36</td>
<td>0.40</td>
<td>0.15</td>
<td>0.19</td>
<td>0.17</td>
<td>0.30</td>
<td>0.26</td>
</tr>
<tr>
<td>T9</td>
<td>-0.02</td>
<td>0.20</td>
<td>0.43</td>
<td>0.61</td>
<td>0.37</td>
<td>0.39</td>
<td>0.38</td>
<td>0.46</td>
<td>0.42</td>
<td>0.59</td>
<td>0.47</td>
<td>0.46</td>
<td>0.41</td>
<td>0.39</td>
<td>0.31</td>
<td>0.17</td>
<td>0.46</td>
<td>0.40</td>
</tr>
<tr>
<td>T10</td>
<td>-0.10</td>
<td>0.17</td>
<td>0.41</td>
<td>0.61</td>
<td>0.43</td>
<td>0.45</td>
<td>0.35</td>
<td>0.51</td>
<td>0.41</td>
<td>0.49</td>
<td>0.52</td>
<td>0.54</td>
<td>0.58</td>
<td>0.41</td>
<td>0.40</td>
<td>0.24</td>
<td>0.43</td>
<td>0.36</td>
</tr>
<tr>
<td>T11</td>
<td>0.14</td>
<td>0.21</td>
<td>0.48</td>
<td>0.61</td>
<td>0.39</td>
<td>0.47</td>
<td>0.35</td>
<td>0.55</td>
<td>0.36</td>
<td>0.53</td>
<td>0.56</td>
<td>0.60</td>
<td>0.53</td>
<td>0.38</td>
<td>0.42</td>
<td>0.22</td>
<td>0.40</td>
<td>0.36</td>
</tr>
<tr>
<td>T12</td>
<td>-0.17</td>
<td>0.27</td>
<td>0.36</td>
<td>0.61</td>
<td>0.33</td>
<td>0.40</td>
<td>0.35</td>
<td>0.29</td>
<td>0.24</td>
<td>0.45</td>
<td>0.52</td>
<td>0.54</td>
<td>0.54</td>
<td>0.50</td>
<td>0.46</td>
<td>0.28</td>
<td>0.38</td>
<td>0.46</td>
</tr>
<tr>
<td>T13</td>
<td>-0.16</td>
<td>0.26</td>
<td>0.42</td>
<td>0.61</td>
<td>0.26</td>
<td>0.30</td>
<td>0.31</td>
<td>0.30</td>
<td>0.18</td>
<td>0.51</td>
<td>0.44</td>
<td>0.45</td>
<td>0.48</td>
<td>0.56</td>
<td>0.31</td>
<td>0.32</td>
<td>0.43</td>
<td>0.39</td>
</tr>
<tr>
<td>T14</td>
<td>0.02</td>
<td>0.23</td>
<td>0.45</td>
<td>0.61</td>
<td>0.38</td>
<td>0.46</td>
<td>0.44</td>
<td>0.51</td>
<td>0.44</td>
<td>0.52</td>
<td>0.56</td>
<td>0.53</td>
<td>0.49</td>
<td>0.40</td>
<td>0.57</td>
<td>0.39</td>
<td>0.36</td>
<td>0.39</td>
</tr>
<tr>
<td>T15</td>
<td>-0.13</td>
<td>0.17</td>
<td>0.20</td>
<td>0.61</td>
<td>0.29</td>
<td>0.27</td>
<td>0.21</td>
<td>0.25</td>
<td>0.29</td>
<td>0.30</td>
<td>0.35</td>
<td>0.33</td>
<td>0.40</td>
<td>0.41</td>
<td>0.42</td>
<td>0.49</td>
<td>0.29</td>
<td>0.31</td>
</tr>
<tr>
<td>T16</td>
<td>-0.13</td>
<td>0.10</td>
<td>0.39</td>
<td>0.61</td>
<td>0.40</td>
<td>0.46</td>
<td>0.45</td>
<td>0.37</td>
<td>0.45</td>
<td>0.54</td>
<td>0.61</td>
<td>0.50</td>
<td>0.47</td>
<td>0.51</td>
<td>0.55</td>
<td>0.48</td>
<td>0.58</td>
<td>0.52</td>
</tr>
<tr>
<td>T17</td>
<td>-0.07</td>
<td>0.28</td>
<td>0.37</td>
<td>0.61</td>
<td>0.29</td>
<td>0.26</td>
<td>0.28</td>
<td>0.29</td>
<td>0.30</td>
<td>0.51</td>
<td>0.44</td>
<td>0.41</td>
<td>0.53</td>
<td>0.40</td>
<td>0.41</td>
<td>0.31</td>
<td>0.45</td>
<td>0.55</td>
</tr>
<tr>
<td>C1</td>
<td>0.00</td>
<td>0.12</td>
<td>-0.01</td>
<td>0.61</td>
<td>0.02</td>
<td>0.10</td>
<td>0.06</td>
<td>0.03</td>
<td>0.04</td>
<td>0.06</td>
<td>-0.11</td>
<td>0.07</td>
<td>0.08</td>
<td>-0.04</td>
<td>0.10</td>
<td>0.06</td>
<td>0.03</td>
<td>-0.02</td>
</tr>
<tr>
<td>C17</td>
<td>0.09</td>
<td>-0.19</td>
<td>-0.08</td>
<td>0.61</td>
<td>-0.20</td>
<td>-0.03</td>
<td>-0.16</td>
<td>0.00</td>
<td>N/A</td>
<td>-0.12</td>
<td>-0.08</td>
<td>-0.06</td>
<td>-0.11</td>
<td>-0.15</td>
<td>-0.07</td>
<td>-0.04</td>
<td>N/A</td>
<td>-0.16</td>
</tr>
</tbody>
</table>
Table 3.1 (Continued).

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
<th>T8</th>
<th>T9</th>
<th>T10</th>
<th>T11</th>
<th>T12</th>
<th>T13</th>
<th>T14</th>
<th>T15</th>
<th>T16</th>
<th>T17</th>
<th>C1</th>
<th>C17</th>
</tr>
</thead>
<tbody>
<tr>
<td>T12</td>
<td>0.55</td>
<td>0.55</td>
<td>0.64</td>
<td>0.60</td>
<td>0.56</td>
<td>0.75</td>
<td>0.70</td>
<td>0.64</td>
<td>0.75</td>
<td>0.87</td>
<td>0.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T13</td>
<td>0.60</td>
<td>0.55</td>
<td>0.56</td>
<td>0.62</td>
<td>0.57</td>
<td>0.64</td>
<td>0.65</td>
<td>0.59</td>
<td>0.75</td>
<td>0.84</td>
<td>0.83</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T14</td>
<td>0.46</td>
<td>0.42</td>
<td>0.62</td>
<td>0.58</td>
<td>0.57</td>
<td>0.71</td>
<td>0.71</td>
<td>0.64</td>
<td>0.76</td>
<td>0.83</td>
<td>0.83</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T15</td>
<td>0.34</td>
<td>0.27</td>
<td>0.49</td>
<td>0.54</td>
<td>0.35</td>
<td>0.43</td>
<td>0.61</td>
<td>0.59</td>
<td>0.60</td>
<td>0.75</td>
<td>0.62</td>
<td>0.73</td>
<td>0.71</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T16</td>
<td>0.40</td>
<td>0.38</td>
<td>0.53</td>
<td>0.59</td>
<td>0.47</td>
<td>0.67</td>
<td>0.62</td>
<td>0.62</td>
<td>0.81</td>
<td>0.79</td>
<td>0.80</td>
<td>0.83</td>
<td>0.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T17</td>
<td>0.40</td>
<td>0.44</td>
<td>0.48</td>
<td>0.63</td>
<td>0.42</td>
<td>0.63</td>
<td>0.59</td>
<td>0.61</td>
<td>0.77</td>
<td>0.76</td>
<td>0.79</td>
<td>0.79</td>
<td>0.74</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>-0.18</td>
<td>-0.01</td>
<td>-0.10</td>
<td>-0.21</td>
<td>0.01</td>
<td>-0.12</td>
<td>-0.17</td>
<td>-0.18</td>
<td>-0.21</td>
<td>-0.17</td>
<td>-0.19</td>
<td>-0.09</td>
<td>-0.14</td>
<td>-0.17</td>
<td>-0.17</td>
<td>-0.20</td>
<td>-0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C17</td>
<td>-0.09</td>
<td>-0.03</td>
<td>-0.14</td>
<td>-0.30</td>
<td>-0.05</td>
<td>-0.07</td>
<td>0.09</td>
<td>0.02</td>
<td>0.03</td>
<td>0.08</td>
<td>0.09</td>
<td>-0.15</td>
<td>-0.23</td>
<td>-0.01</td>
<td>0.01</td>
<td>N/A</td>
<td>-0.22</td>
<td>-0.07</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Tx = Treatment Condition (modular = 1, standard = 0); B = Brief Problem Checklist score, where B1 is the score at week 1 of treatment, B2 is the score at week 2 etc.; T = Top Problems Assessment score at week 1 of treatment, T2 is the score at week 2 etc.; C1 = covariate indicating the number of weeks that first weekly assessment came after first treatment session; C17 = covariate indicating the number of weeks that the last assessment up through week 17 came after the preceding treatment session; N/A = not applicable because C17 = 1 for all values of the other variable, thus there was no variance in C17 to compute a correlation with the other variable.
Univariate Latent Change Score Models

Table 3.2 shows the fit statistics for the univariate latent change score models fitted for the symptoms and top problems measures. Good absolute model fit would be evidenced by a nonsignificant $\chi^2 (p > .05)$, RMSEA < .05, lower bound of 90% CI = 0, upper bound of 90% CI < .10, CFI $\geq$ .95, and SRMR $\leq$ .08 (see, e.g., Kline, 2011). None of the models fit either measure well, though the four models with dynamic noise showed marginally acceptable fit for both measures, with RMSEA < .08, upper bound of 90% CI < .10, CFI $\geq$ .90, and SRMR $\leq$ .08. However, three of the dynamic noise models (constant change, linear change, and linear change plus dual change) had estimation problems; the latent covariance (psi) matrix was nonpositive definite due to negative variance of one or two growth factors. Fixing the variance of the growth factor(s) to zero did not improve model fit appreciably, and in some cases led to worse fit. The dual change plus dynamic noise model had the best absolute fit to the data and did not involve estimation problems for both measures. The dual change plus dynamic noise model also had the best relative fit compared to nested models for both measures. The chi-square difference test with scaling correction factor applied for MLR estimation showed that the dual change plus dynamic noise model fit the data significantly better than the no change, constant change, proportional change, and dual change models (all $p$s < .01). On the other hand, adding a linear change component to the dual change plus dynamic noise model did not significantly improve fit ($p$s > .05). The dual change plus dynamic noise model also had one of the lowest AIC and BIC values compared to the other models. Thus the dual change plus dynamic noise models for symptoms and top problems were combined in bivariate LCS models.
### Table 3.2. Fit Statistics for Univariate Latent Change Score Models for Symptoms and Top Problems

<table>
<thead>
<tr>
<th></th>
<th>Npar</th>
<th>LL</th>
<th>LL SCF</th>
<th>AIC</th>
<th>BIC</th>
<th>χ²</th>
<th>df</th>
<th>P</th>
<th>χ²</th>
<th>p for Δχ²</th>
<th>RMSEA</th>
<th>90% CI</th>
<th>CFI</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symptoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Change</td>
<td>3</td>
<td>-3579</td>
<td>3.234</td>
<td>7164</td>
<td>7172</td>
<td>446.9</td>
<td>167</td>
<td>&lt;.0002</td>
<td>1.218</td>
<td>&lt;.0001^a</td>
<td>0.118</td>
<td>[0.105, 0.131]</td>
<td>0.739</td>
<td>0.107</td>
</tr>
<tr>
<td>Constant Change</td>
<td>6</td>
<td>-3475</td>
<td>2.297</td>
<td>6962</td>
<td>6978</td>
<td>277.0</td>
<td>164</td>
<td>&lt;.0001</td>
<td>1.216</td>
<td>.0043^a</td>
<td>0.076</td>
<td>[0.060, 0.091]</td>
<td>0.894</td>
<td>0.077</td>
</tr>
<tr>
<td>Proportional Change</td>
<td>4</td>
<td>-3541</td>
<td>3.572</td>
<td>7091</td>
<td>7102</td>
<td>392.3</td>
<td>166</td>
<td>&lt;.0001</td>
<td>1.198</td>
<td>&lt;.0001^a</td>
<td>0.107</td>
<td>[0.093, 0.120]</td>
<td>0.583</td>
<td>0.113</td>
</tr>
<tr>
<td>Dual Change</td>
<td>7</td>
<td>-3473</td>
<td>2.255</td>
<td>6960</td>
<td>6980</td>
<td>275.2</td>
<td>163</td>
<td>&lt;.0001</td>
<td>1.211</td>
<td>.0069^a</td>
<td>0.076</td>
<td>[0.060, 0.091]</td>
<td>0.895</td>
<td>0.074</td>
</tr>
<tr>
<td>Linear Change</td>
<td>10</td>
<td>-3471</td>
<td>1.917</td>
<td>6963</td>
<td>6991</td>
<td>271.9</td>
<td>160</td>
<td>&lt;.0001</td>
<td>1.213</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear Change +</td>
<td>11</td>
<td>-3471</td>
<td>2.026</td>
<td>6964</td>
<td>6994</td>
<td>273.9</td>
<td>159</td>
<td>&lt;.0001</td>
<td>1.201</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant Change +</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic Noise</td>
<td>7</td>
<td>-3460</td>
<td>2.444</td>
<td>6935</td>
<td>6954</td>
<td>255.9</td>
<td>163</td>
<td>&lt;.0001</td>
<td>1.203</td>
<td></td>
<td></td>
<td>[0.052, 0.085]</td>
<td>0.913</td>
<td>0.072</td>
</tr>
<tr>
<td>Dual Change +</td>
<td>8</td>
<td>-3459</td>
<td>2.465</td>
<td>6933</td>
<td>6956</td>
<td>255.0</td>
<td>162</td>
<td>&lt;.0001</td>
<td>1.194</td>
<td></td>
<td></td>
<td>[0.052, 0.085]</td>
<td>0.913</td>
<td>0.070</td>
</tr>
<tr>
<td>Linear Change +</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic Noise</td>
<td>11</td>
<td>-3455</td>
<td>1.955</td>
<td>6932</td>
<td>6962</td>
<td>246.2</td>
<td>159</td>
<td>&lt;.0001</td>
<td>1.205</td>
<td></td>
<td></td>
<td>[0.050, 0.084]</td>
<td>0.919</td>
<td>0.070</td>
</tr>
<tr>
<td>Linear Change +</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dual Change +</td>
<td>12</td>
<td>-3455</td>
<td>1.961</td>
<td>6933</td>
<td>6967</td>
<td>247.0</td>
<td>158</td>
<td>&lt;.0001</td>
<td>1.200</td>
<td>.0774^a</td>
<td>0.069</td>
<td>[0.052, 0.085]</td>
<td>0.917</td>
<td>0.069</td>
</tr>
<tr>
<td><strong>Top Problems</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Change</td>
<td>3</td>
<td>-2781</td>
<td>2.676</td>
<td>5568</td>
<td>5577</td>
<td>707.7</td>
<td>167</td>
<td>&lt;.0001</td>
<td>1.182</td>
<td>&lt;.0001^a</td>
<td>0.164</td>
<td>[ .152, .177]</td>
<td>0.551</td>
<td>0.182</td>
</tr>
<tr>
<td>Constant Change</td>
<td>6</td>
<td>-2553</td>
<td>1.826</td>
<td>5119</td>
<td>5135</td>
<td>321.2</td>
<td>164</td>
<td>&lt;.0001</td>
<td>1.186</td>
<td>&lt;.0001^a</td>
<td>0.089</td>
<td>[ .075, .104]</td>
<td>0.869</td>
<td>0.084</td>
</tr>
<tr>
<td>Proportional Change</td>
<td>4</td>
<td>-2690</td>
<td>2.847</td>
<td>5388</td>
<td>5399</td>
<td>559.8</td>
<td>166</td>
<td>&lt;.0001</td>
<td>1.169</td>
<td>&lt;.0001^a</td>
<td>0.141</td>
<td>[ .128, .154]</td>
<td>0.673</td>
<td>0.182</td>
</tr>
<tr>
<td>Dual Change</td>
<td>7</td>
<td>-2552</td>
<td>1.827</td>
<td>5118</td>
<td>5137</td>
<td>319.7</td>
<td>163</td>
<td>&lt;.0001</td>
<td>1.182</td>
<td>&lt;.0001^a</td>
<td>0.09</td>
<td>[ .075, .104]</td>
<td>0.870</td>
<td>0.081</td>
</tr>
<tr>
<td>Linear Change</td>
<td>10</td>
<td>-2540</td>
<td>1.605</td>
<td>5100</td>
<td>5128</td>
<td>299.7</td>
<td>160</td>
<td>&lt;.0001</td>
<td>1.184</td>
<td></td>
<td></td>
<td>[ .070, .100]</td>
<td>0.884</td>
<td>0.076</td>
</tr>
<tr>
<td>Linear Change +</td>
<td>11</td>
<td>-3471</td>
<td>2.026</td>
<td>6964</td>
<td>6994</td>
<td>273.9</td>
<td>159</td>
<td>&lt;.0001</td>
<td>1.201</td>
<td></td>
<td></td>
<td>[ .062, .093]</td>
<td>0.893</td>
<td>0.073</td>
</tr>
<tr>
<td>Constant Change +</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic Noise</td>
<td>7</td>
<td>-2524</td>
<td>1.868</td>
<td>5062</td>
<td>5082</td>
<td>273.4</td>
<td>163</td>
<td>&lt;.0001</td>
<td>1.181</td>
<td></td>
<td></td>
<td>[ .059, .090]</td>
<td>0.908</td>
<td>0.076</td>
</tr>
</tbody>
</table>
Table 3.2 (Continued).

<table>
<thead>
<tr>
<th>Model Description</th>
<th>Npar</th>
<th>Log Likelihood</th>
<th>AIC</th>
<th>BIC</th>
<th>Deviance Difference</th>
<th>p</th>
<th>Scaling Correction Factor LL SCF</th>
<th>Npar</th>
<th>Log Likelihood</th>
<th>AIC</th>
<th>BIC</th>
<th>Deviance Difference</th>
<th>p</th>
<th>Scaling Correction Factor LL SCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dual Change + Dynamic Noise</td>
<td>8</td>
<td>-2.524</td>
<td>1.750</td>
<td>5064</td>
<td>5086</td>
<td>162 &lt;.0001</td>
<td>1.182</td>
<td>0.075</td>
<td>[.060, .091]</td>
<td>0.908</td>
<td>0.073</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear Change + Dynamic Noise^</td>
<td>11</td>
<td>-2.520</td>
<td>1.577</td>
<td>5063</td>
<td>5094</td>
<td>266.5 159 &lt;.0001</td>
<td>1.183</td>
<td>0.075</td>
<td>[.059, .091]</td>
<td>0.911</td>
<td>0.072</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear Change + Dual Change + Dynamic Noise^</td>
<td>12</td>
<td>-2.520</td>
<td>1.529</td>
<td>5065</td>
<td>5098</td>
<td>266.2 158 &lt;.0001</td>
<td>1.185 .1803^</td>
<td>0.076</td>
<td>[.056, .091]</td>
<td>0.910</td>
<td>0.072</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Npar = number of parameters estimated, LL = log likelihood, LL SCF = scaling correction factor for log likelihood when robust maximum likelihood estimation was used, AIC = Akaike’s information criteria, BIC = Bayesian information criteria, $\chi^2$ SCF = scaling correction factor for chi-square when robust maximum likelihood estimation was used, $p$ for $\Delta\chi^2$ = significance level of chi-square difference test, CFI = comparative fit index, RMSEA = root mean square error of approximation, 90% CI RSMSEA = 90% confidence intervals of the root mean square error of approximation, SRMR = standardized root mean square residual.

^These models had estimation problems, the latent covariance (psi) matrix was nonpositive definite due to negative variance of the constant change and/or linear change growth factors. ^Obtained by comparing to Dual Change + Dynamic Noise model.
Bivariate Latent Change Score Models

Table 3.3 shows the fit statistics and Table 3.4 shows the parameter estimates and standard errors for the bivariate dual change plus dynamic noise models. Due to low covariance coverage (i.e., missing data resulting in a low proportion of cases contributing data to some variances or covariances), some fit statistics could not be computed when MLR estimation was used. Thus models using maximum likelihood (ML) estimation were used to generate the fit statistics, but the models using MLR estimation were used to generate the parameter estimates, standard errors, and $p$ values for all results henceforth.

For all four bivariate models, only the SRMR indicated acceptable fit, the other absolute fit indices suggested inadequate fit. To assess relative fit, the unidirectional models were compared to the no coupling model. The unidirectional model with coupling effects from top problems to symptom change did not fit significantly better than the no coupling model ($p > .05$). On the other hand, the unidirectional model with coupling effects from symptoms to change in top problems fit significantly better than the no coupling model ($p < .001$). Adding coupling effects bidirectionally improved fit marginally ($p = .055$). Examining the coupling parameters in the four bivariate models, only the coupling effect going from symptoms to change in top problems was significant ($p < .05$), the coupling effect in the reverse direction was not significant. These results suggest that symptoms may be a leading indicator of change in top problems severity, and that the most appropriate model is the unidirectional model with a coupling effect from symptoms to change in top problem severity. However, these findings should be considered preliminary due to the inadequate absolute fit of the models.

---

9 Attempts to improve the absolute fit of the bivariate models were unsuccessful, including freeing the coupling parameters that were constrained to be equal, reducing the number of weeks in the model to improve covariance coverage, and removing the dynamic noise component.
### Table 3.3. Fit Statistics of the Bivariate Dual Change + Dynamic Noise Models

<table>
<thead>
<tr>
<th></th>
<th>Npar</th>
<th>LL</th>
<th>AIC</th>
<th>BIC</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$</th>
<th>$p$ for $\Delta \chi^2$</th>
<th>RMSEA</th>
<th>90% CI</th>
<th>CFI</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Coupling$^\wedge$</td>
<td>21</td>
<td>-5762</td>
<td>11566</td>
<td>11625</td>
<td>1607.1</td>
<td>608</td>
<td>&lt;.0001</td>
<td>0.117</td>
<td>[0.110, 0.124]</td>
<td>0.762</td>
<td>0.073</td>
<td></td>
</tr>
<tr>
<td>Unidirectional Coupling T $\rightarrow$ B</td>
<td>22</td>
<td>-5762</td>
<td>11568</td>
<td>11629</td>
<td>1606.8</td>
<td>607</td>
<td>&lt;.0001</td>
<td>0.597$^a$</td>
<td>0.117</td>
<td>[0.110, 0.124]</td>
<td>0.762</td>
<td>0.074</td>
</tr>
<tr>
<td>Unidirectional Coupling B $\rightarrow$ T</td>
<td>22</td>
<td>-5757</td>
<td>11557</td>
<td>11618</td>
<td>1596.0</td>
<td>607</td>
<td>&lt;.0001</td>
<td>0.001$^a$</td>
<td>0.117</td>
<td>[0.110, 0.124]</td>
<td>0.764</td>
<td>0.071</td>
</tr>
<tr>
<td>Bidirectional Coupling</td>
<td>23</td>
<td>-5755</td>
<td>11555</td>
<td>11620</td>
<td>1592.3</td>
<td>606</td>
<td>&lt;.0001</td>
<td>0.055$^b$</td>
<td>0.116</td>
<td>[0.110, 0.123]</td>
<td>0.765</td>
<td>0.072</td>
</tr>
</tbody>
</table>

**Note.** B = Brief Problem Checklist, T = Top Problems Assessment, Npar = number of parameters estimated, LL = log likelihood, AIC = Akaike’s information criteria, BIC = Bayesian information criteria, $p$ for $\Delta \chi^2$ = significance level of chi-square difference test, RMSEA = root mean square error of approximation, 90% CI RMSEA = 90% confidence intervals of the root mean square error of approximation, SRMR = standardized root mean square residual.

$^\wedge$This model had estimation problems, the latent covariance (psi) matrix was nonpositive definite. $^a$Obtained by comparing to No Coupling model. $^b$Obtained by comparing to Unidirectional Coupling B $\rightarrow$ T model.
Table 3.4. *Parameter Estimates and Standard Errors of the Bivariate Dual Change + Dynamic Noise Models*

<table>
<thead>
<tr>
<th></th>
<th>No Coupling</th>
<th>Unidirectional Coupling T → B</th>
<th>Unidirectional Coupling B → T</th>
<th>Bidirectional Coupling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>9.29 (0.46)***</td>
<td>9.31 (0.45)***</td>
<td>9.31 (0.44)***</td>
<td>9.26 (0.44)***</td>
</tr>
<tr>
<td>T</td>
<td>6.72 (0.20)***</td>
<td>6.73 (0.19)***</td>
<td>6.76 (0.20)***</td>
<td>6.73 (0.20)***</td>
</tr>
<tr>
<td>Constant Change</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.38 (0.42)</td>
<td>0.54 (0.47)</td>
<td>0.34 (0.36)</td>
<td>0.31 (0.42)</td>
</tr>
<tr>
<td>T</td>
<td>-0.032 (0.23)</td>
<td>0.02 (0.18)</td>
<td>0.05 (0.15)</td>
<td>-0.03 (0.16)</td>
</tr>
<tr>
<td>Proportional Change</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>-0.07 (0.05)</td>
<td>-0.17 (0.21)</td>
<td>-0.07 (0.05)</td>
<td>-0.17 (0.15)</td>
</tr>
<tr>
<td>T</td>
<td>-0.02 (0.04)</td>
<td>-0.03 (0.03)</td>
<td>-0.15 (0.06)**</td>
<td>-0.14 (0.05)**</td>
</tr>
<tr>
<td>Coupling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T → B</td>
<td></td>
<td>0.11 (0.23)</td>
<td></td>
<td>0.15 (0.16)</td>
</tr>
<tr>
<td>B → T</td>
<td></td>
<td></td>
<td>0.09 (0.04)*</td>
<td>0.09 (0.03)*</td>
</tr>
</tbody>
</table>

*Note.* B = Brief Problem Checklist, T = Top Problems Assessment.

*p < .05, **p < .01, ***p < .001.

In the unidirectional coupling model from symptoms to top problems, the intercepts indicate that the change trajectories of symptoms and top problems latent true scores started at 9.31 and 6.76 respectively in week 1. The shape of change depends on the constant change and proportional change components. The nonsignificant positive constant change component of both measures suggests that there is a small element of worsening in symptoms by 0.34 points, and in top problems by 0.05 points, at a constant rate each week. However, this constant change is reversed by the negative proportional change component indicating weekly decreases in 7% of the prior week’s estimated symptom score and in 15% of the prior week’s estimated top problems score, resulting in a net decrease in scores over time. In addition, the proportional change effect is self-limiting, because the magnitude of change from the previous week is proportional to the score of the previous week. For example, the week 1 estimated mean
symptom score was 9.31, thus the mean change going from week 1 to week 2 = 0.34 + (-
.07*9.31) = -0.31, a net decrease in symptom severity. The week 2 mean symptom score was
9.31 – 0.31 = 9.00, thus the mean change going from week 2 to week 3 = 0.34 + (-.07*9.00) = -
0.29, a smaller net decrease in symptom severity than the previous week. Scores decreased at a
slower rate with each passing week, resulting in a declining exponential curve. The significant
positive coupling effect from symptoms to top problems suggest that higher symptoms in one
week preceded an increase in top problem severity the next week, or conversely, lower
symptoms preceded a decrease in top problem severity.

The variances of the latent scores and change components, and their correlations with one
another, indicate between-youth differences. The variances of the latent true scores at week 1
were significantly greater than zero for each measure (symptoms = 18.6, top problems = 3.96, ps
< .001), reflecting substantial variability among youths in their severity of symptoms and top
problems in week 1. The variance of the constant change component was not significantly
different from zero for either measure (symptoms = 0.10, top problems = 0.11, ps > .05),
reflecting low degree of variability among youths in this aspect of change. Week 1 latent true
score of symptoms was significantly correlated with the week 1 latent true score (r = .50, p <
.001) and constant change component (r = -.50, p < .001) of top problems, suggesting that youths
who had more severe symptoms at week 1 also had more severely rated top problems at week 1,
but less weekly worsening of top problems. The remaining correlations between latent variables
were not significant (all ps < .05), including correlations between the week 1 latent true score
and constant change component of symptoms, between the week 1 latent true scores of top
problems and constant change components of both measures, and between the two constant
change components.
Models with Covariates and Treatment Condition Added

After selecting the bivariate model with unidirectional coupling from symptoms to top problem change, the two covariates were added to the model, with the intercept and constant change component of the symptoms and top problems measures regressed on each centered covariate (see Dogan, Stockdale, Widaman, & Conger, 2010). This model generated a nonpositive definite first-order derivative product matrix due to issues with the path from the last treatment-assessment discrepancy covariate and the top problems constant change component. The last treatment-assessment discrepancy covariate was removed from the model because of the estimation issue and because it was not significantly associated with the intercept or constant change components of either measure (all $p$s > .05). The first treatment–assessment discrepancy covariate was significantly associated with the top problems constant change component ($B = -0.030, p < .05$), thus it was retained in the model. There was no change in relative fit whether the intercept and constant change component of the symptom measure were regressed on the first treatment–assessment discrepancy covariate or not ($\Delta \chi^2 = 0.064, \Delta df = 2, p > .05$). In addition, absolute fit indices were nearly identical, so these paths were pruned from the model, leaving only the intercept and constant change component of top problems to be covaried by first treatment–assessment discrepancy (in the more restrictive model, $\chi^2 = 1636, df = 639, p < .0001$, RMSEA = .114, 90% CI [.107, .121], CFI = .763, SRMR = .075).

Next, treatment condition was added to the model, and three models were fitted to assess treatment effects on top problems as the outcome variable and symptom severity as the mediator variable, given that symptom severity was shown to be a possible leading indicator of change in top problems: (a) only direct effects were estimated by regressing the intercept and the weekly latent change scores of top problems on treatment condition and constraining the parameter
estimates to be equal across weeks (path \( c \) of mediation model, also considered the total effect of treatment condition due in the absence of indirect effect); (b) only indirect effects were estimated by regressing the intercept and constant change component of symptom severity on treatment condition (path \( a \) of mediation model; the unidirectional coupling parameter regressing changes in top problems on symptom severity serves as path \( b \) of mediation model); and (c) both direct and indirect effects were estimated. Table 3.5 shows that all three models have absolute fit indices that are highly similar to one another and to the previous set of bivariate models without the inclusion of treatment condition and the covariate. However, relative fit was significantly better in the model estimating both direct and indirect effects compared to the model estimating only direct effects. The models estimating only indirect effects did not fit the data significantly worse than the model estimating both direct and indirect effects, suggesting that the indirect effects only model is the most parsimonious model among the three. However, indirect effects, which differed across weeks, were not significant in either the indirect and direct effects model (e.g., weeks 5, 9, 13, and 17 \( B = 0.046 - 0.079 \), all \( ps > .05 \)) or the indirect effects only model (e.g., weeks 5, 9, 13, and 17 \( B = -0.040 - 0.023 \), all \( ps > .05 \)). The direct effect, constrained to be equal across weeks, was not significant in either the indirect and direct effects model (\( B = -0.108, p > .05 \)) or the direct effects only model (\( B = -0.035, p > .05 \)). Interestingly, in the indirect and direct effects model, total effects of treatment on latent change in top problem severity was significant in later weeks, with MATCH leading to greater reduction in top problem severity (e.g., week 5 \( B = -0.029 \), week 9 \( B = -0.057, ps > .05 \); week 13 \( B = -0.061 \), week 17 \( B = -0.056, ps < .05 \)). Based on relative fit, the indirect effects only model was chosen as the model to examine more closely for potential effects of treatment condition on changes in symptom and top problem severity.
Table 3.5. *Fit Statistics of the Bivariate Dual Change + Dynamic Noise Models With and Without Indirect Effect and Direct Effect of Treatment Condition Estimated, Covarying Number of Weeks Between First Treatment Session and First Assessment*

<table>
<thead>
<tr>
<th></th>
<th>Npar</th>
<th>LL</th>
<th>AIC</th>
<th>BIC</th>
<th>( \chi^2 )</th>
<th>Df</th>
<th>( p )</th>
<th>( p ) for ( \Delta \chi^2 )</th>
<th>RMSEA</th>
<th>90% CI</th>
<th>CFI</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Effects Only</td>
<td>26</td>
<td>-5753</td>
<td>11558</td>
<td>11631</td>
<td>1703.3</td>
<td>671</td>
<td>&lt;.0001</td>
<td>0.015</td>
<td>0.113</td>
<td>[.107, .120]</td>
<td>0.757</td>
<td>0.080</td>
</tr>
<tr>
<td>Indirect Effects Only</td>
<td>26</td>
<td>-5751</td>
<td>11554</td>
<td>11626</td>
<td>1698.5</td>
<td>671</td>
<td>&lt;.0001</td>
<td>0.165</td>
<td>0.113</td>
<td>[.106, .120]</td>
<td>0.758</td>
<td>0.078</td>
</tr>
<tr>
<td>Both Direct and Indirect Effects</td>
<td>28</td>
<td>-5749</td>
<td>11554</td>
<td>11632</td>
<td>1694.9</td>
<td>669</td>
<td>&lt;.0001</td>
<td>0.113</td>
<td>[.106, .120]</td>
<td>0.758</td>
<td>0.077</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Npar = number of parameters estimated, LL = log likelihood, AIC = Akaike’s information criteria, BIC = Bayesian information criteria, \( p \) for \( \Delta \chi^2 \) = significance level of chi-square difference test, RMSEA = root mean square error of approximation, 90% CI RSMSEA = 90% confidence intervals of the root mean square error of approximation, SRMR = standardized root mean square residual.

\(^a\) Obtained by comparing to model with Both Direct and Indirect Effects.
Table 3.6. *Parameter Estimates and Standard Errors of the Bivariate Dual Change + Dynamic Noise Model With Indirect Effect of Treatment Condition Estimated, Covarying Number of Weeks Between First Treatment Session and First Assessment*

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Variance</th>
<th>Association with Treatment Condition</th>
<th>Association with Covariate</th>
<th>Correlation with Intercept of B</th>
<th>Correlation with Intercept of T</th>
<th>Correlation with Constant Change of B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>8.76 (0.49)***</td>
<td>17.7 (2.93)***</td>
<td>1.09 (0.69)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>6.77 (0.20)***</td>
<td>4.01 (0.64)***</td>
<td>&lt;0.001 (0.062)</td>
<td>.49 (.08)***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant Change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.37 (0.35)</td>
<td>0.10 (0.11)</td>
<td>-0.10 (.08)</td>
<td>.48 (.27)</td>
<td>.18 (.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>0.05 (0.16)</td>
<td>-0.12 (.08)</td>
<td>-0.03 (0.01)*</td>
<td>-0.51 (0.18)***</td>
<td>.17 (.16)</td>
<td>-.31 (.25)</td>
<td></td>
</tr>
<tr>
<td>Proportional Change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>-0.06 (0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>-0.16 (0.06)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coupling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B → T</td>
<td>0.09 (0.04)*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indirect Effect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 5</td>
<td>0.02 (0.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 9</td>
<td>-0.02 (0.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 13</td>
<td>-0.04 (0.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 17</td>
<td>-0.04 (0.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* B = Brief Problem Checklist, T = Top Problems Assessment.

*p < .05, **p < .01, ***p < .001.
Table 3.6 shows that treatment condition had no significant effect on the week 1 latent true score and constant change component of symptom severity, consistent with the lack of a significant indirect effect on change in top problem severity via symptom reduction. The first assessment-treatment discrepancy covariate remained significantly and negatively associated with the top problems constant change component. The mean intercepts, constant change components, proportional change components, and coupling effect in this model are similar in magnitude, direction, and significance level to those in the previous bivariate model without treatment condition and covariate added. The variances of the week 1 latent true scores and change components, and their correlations with one another, were also similar to those in the previous bivariate model. Thus their interpretation would be similar to that described in the previous section.

**Discussion**

Using data from the Weisz et al. (2012) effectiveness trial, the present study investigated whether personalized treatment goal achievement was a change process underlying the superior effects of modular over standard EBPs in ameliorating internalizing and externalizing symptoms in a sample of treatment-seeking youths. Findings are tentative due to difficulties with model fit to the bivariate data. With that strong caveat in mind, the present study results did not support parent-nominated achievement of personalized treatment goals as a candidate change process contributing to parent-reported treatment gains of modular over standard EBPs. A logical criterion for a change process is that it precedes symptom change (e.g., Doss, 2004), but changes in top problem severity appeared to follow symptom change in the present sample. In addition, symptom change did not significantly mediate the effects of treatment condition on top problem severity; instead, it appeared to drive top problem improvement across modular and standard
EBPs. This preliminary finding is consistent with the interpretation that change processes and mechanisms other than improvement in parent-nominated treatment goals are responsible for the differential effectiveness of modular vs. standard EBPs, and that overall symptom reduction may either equip youths to achieve these specific treatment goals, especially if the top problems were broader than specific internalizing or externalizing symptoms, or convey to parents that these goals are being met.

These findings and their interpretation may have been complicated by the particular measures used in the present study, the TPA and the BPC. Both measures served multiple functions in the trial—they were not only used as outcome measures for research, but as part of a measurement feedback system to provide therapists and their supervisors with information about the youths’ treatment progress and facilitate clinical decision-making (see Bickman, 2008). In the Weisz et al. (2012) trial, the TPA and BPC scores and other information (e.g., therapy content and practices used during sessions) were entered weekly into a website by research assistants; these data were then displayed on an internet-based “clinical dashboard” (Chorpita et al., 2008) for supervisors to review. Presumably, the TPA and BPC data were both used by therapists to guide their treatment focus, and it is unclear how exactly they did so, for example, if they prioritized one over the other in making treatment decisions. The bivariate LCS model appears to have been developed for, and used predominantly, to study developmental processes that unfold over time (e.g., Dogan et al., 2010; McArdle, Hamagami, Chang, & Hishinuma, 2014; see also Grimm et al., 2017). It is unclear how an intervention with a built-in feedback process would affect model assumptions and performance. Although bivariate LCS modeling has been used with a small number of intervention studies (e.g., Marker, Comer, Abramova, & Kendall, 2013; Radkovsvky, McArdle, Bockting, & Berking, 2014), the key variables were not
explicitly used to inform intervention in those studies. In addition, between-treatment differences in addressing personalized treatment goals might have been attenuated because the TPA was used by therapists in both treatment conditions to inform treatment planning.

A related measurement issue has to do with informant discrepancies, which are common in youth psychotherapy research (De Los Reyes, Thomas, Goodman, & Kundey, 2013). Youths and parents identified their top problems separately, and therapists received feedback from both informants. In some cases, therapists worked primarily with youths (usually when internalizing symptoms were primary), and in other cases, therapists worked primarily with parents (usually when externalizing symptoms were primary). It is unclear how therapists chose to integrate information about different top problems identified by parent vs. youth, or different severity ratings on the same top problems across informants. It is plausible that therapists are predominantly focused on addressing youth-reported top problems, especially with youths who have primarily internalizing symptoms, leading to overall symptom change, which in turn heightened parent awareness of their identified treatment goals being met. Unfortunately, due to challenges with modeling the functional form of the youth-reported data, the relationships among youth-identified top problem, parent-identified top problems, and symptoms reported by both informants, could not be clarified.

In addition, the relationships between top problem and symptom severity may have been more complex than those modeled in the present study. There are other variations of bivariate LCS models that were not used, including certain nonlinear LCS models and a “changes-to-changes” model that assessed coupling effects of changes in one variable on changes in the other variable (see Grimm et al., 2017). In additional, multiple-group bivariate LCS models (see McArdle et al., 2014), in which change components and coupling effects are compared across
groups, were not fitted. These models can detect potential differences in the direction and magnitude of relationships between treatment conditions, or subgroups of youths (e.g., those with primarily internalizing vs. externalizing symptoms). Some modifications of bivariate LCS models were attempted in which equality constraints across time were relaxed due to the possibility of coupling effects differing as a function of time (see Dogan et al., 2012), but limits on computational power precluded those analyses. Next steps may include exploring additional bivariate LCS models that may fit the data better, especially considering the model fit issues in the present study.

**Limitations, Strengths, and Future Directions**

The major limitation of the present study is that the models tested simply did not represent the observed data very well, despite extensive efforts to identify better fitting models. Virtually every type of longitudinal growth model in the textbook (see Grimm et al., 2017) was fitted to the data and only marginally acceptable fit was achieved for the univariate LCS models, which declined in the bivariate LCS models. A promising future direction could be to analyze the present study data using dynamic structural equation models (DSEM), a cutting edge class of methods developed for intensive longitudinal data (Asparouhov, Hamaker, & Muthén, 2017; Hamaker, Asparouhov, Brose, Schmiedek, & Muthén, 2017). DSEM allows a time series model to be specified at the within-person level, with individual differences estimated for between-persons parameters (Hamaker et al., 2017), and has recently been incorporated into SEM software (Mplus version 8, Muthén & Muthén, 2017). DSEM might be what is needed to accurately model the dynamics between treatment outcomes and change processes.

Another limitation is that only one candidate change process was examined. Researchers recommend assessing multiple candidate change processes and mechanisms, including those not
associated with specific hypotheses, to establish the specificity of significant mediators as
candidate processes and mechanisms in relation to nonmediating variables (Doss, 2004; Kazdin,
2007). There are certainly other candidate change processes and mechanisms that would be
worth exploring. For example, modular EBPs may be better at engaging both therapists and
families in treatment than standard EBPs. A survey of the therapists in the present study showed
that they had more positive attitudes towards EBPs that were not organized in a typical
manualized fashion (Borntrager et al., 2009). It remains to be tested if those positive attitudes
would translate into better engagement in learning MATCH and higher competency in using it.
Also, modular EBPs may allow youths to receive more instruction or practice in core modules
that are expected to contain the active ingredients, because the decision-making flowcharts
explicitly guide therapists to repeat or return to core modules until sufficient progress is made.
One study with phobic adults found that a standardized EBP outperformed an individually-
tailored version of the EBP because the standardized version included greater usage of exposure,
the core therapy process thought to cause symptom improvement (Schulte, Künzel, Pepping, &

The present study also has several strengths. First, this one of the few studies of change
processes in the context of an effectiveness trial. The majority of change process research has
been conducted under the highly controlled conditions of efficacy trials. Studying how
psychotherapies work in everyday clinical service contexts can have considerable value if the
methods used are rigorous. Most psychotherapies are ultimately intended to be implemented by
practitioners, treating referred individuals, in clinical service settings; therefore findings in
effectiveness trials have greater potential to generalize to those conditions (Weisz et al., 2014).
Second, this is one of the few psychotherapy studies (see Marker et al., 2013) that aimed to
disentangle temporal relationships between change processes and outcomes, had sufficiently frequent measurements to do so, and used quantitative methods designed to assess the direction of variable relationships. Progress in understanding how psychotherapy works has long been stymied by infrequent measurement schedules that do not allow temporal precedence of candidate change processes and mechanisms over outcomes to be established—an important step towards understanding whether the candidate change processes or mechanism are in fact causing improvement in outcome (e.g. Doss, 2004; Kazdin, 2007). Finally, the present study may serve as an example of the potential utility of advanced statistical approaches for solving problems in studying therapeutic change—as well as some of the challenges involved—thereby informing the selection of statistical methods to answer questions about how psychotherapies work.
References


**General Discussion**

I proposed this dissertation because I was intrigued by the idea that the field of psychotherapy research has not clearly figured out how exactly psychotherapies work—despite hundreds or thousands of psychotherapy trials that have been conducted, and the mountain of evidence they have generated to demonstrate that psychotherapy works (e.g., Smith, Glass, & Miller, 1980; Weisz et al., 2017). There is still debate among researchers and clinicians about the change processes and mechanisms through which psychotherapies work (see, e.g., Lilienfeld & Arkowitz, 2012). The debate rages on in part due to methodological limitations and mixed findings of existing studies (see Crits-Christoph, Connolly Gibson, & Mukherjee, 2013; Kazdin, 2007; Weisz, Ng, Rutt, Lau, & Masland, 2013). As a budding youth psychotherapy researcher, I was eager to contribute to knowledge about how psychotherapies work to alleviate common emotional and behavioral problems among youths, particularly the evidence-based psychotherapies (EBPs) shown to be efficacious in multiple randomized controlled trials (RCTs). My aims were to identify mediators and nonmediators of the effects of EBPs for youth depression to inform selection of promising candidates for further study as change processes and mechanisms, while applying advanced quantitative methods with the potential to address some limitations of the extant research and facilitate the identification of treatment mediators. In this general discussion of the three dissertation studies, I evaluate the extent to which my aims were achieved, focusing on common strengths and limitations, and concluding with some ideas for a future research agenda.

**Identifying Mediators and Nonmediators of Treatment Outcome**

Among the three studies, only Study 1 identified mediators that may be change mechanisms specific to EBPs. Negative thinking, and social skills and relations, significantly
mediated the effects of cognitive behavioral therapy (CBT), behavioral therapy (BT), and interpersonal therapy (IPT) compared to control conditions across 24 and 12 RCTs, including more than 2,000 and 1,000 depressed youths, respectively. Negative thinking and social skills and relations could reasonably be considered robust mediators, considering that the optimal sample size for meta-analytic structural equation modeling (MASEM) using the two-stage structural equation modeling (TSSEM) approach is at least 1,000 participants across 10 studies (Cheung & Chan, 2005). Though long hypothesized as a change mechanism of CBT for youth depression, individual mediation studies of negative thinking suffer from a small, nonrepresentative sample and conflicting results. Thus Study 1 may have produced the most conclusive evidence to date that negative thinking is a robust mediator of EBPs for youth depression. Moreover, social skills and relations did not significantly mediate treatment effects in the two individual studies that tested it, but nevertheless emerged as a robust mediator in Study 1. These results may signal that there are sufficient RCTs assessing these two mediators at pre- and posttreatment, and that the next steps could include assessing these mediators during treatment, multiple times if possible, to enable analysis of whether change in these robust mediators preceded change in the outcome. Other research directions would be to compare EBP conditions that vary the extent to which negative thinking or social skills and relations are targeted, and assess whether this experimental manipulation results in greater change in the mediator as well as outcome.

There was one other mediator identified: youth-reported secondary control coping. However, the pattern of results in Study 2 suggests that this mediator may be a change mechanism of usual care (UC) rather than that of CBT. Given the heterogeneous nature of UC, future research may zoom in on the aspects of UC that might or might not be improving
secondary control coping, whether in the context of a RCT or not. It may be prudent to hold off on rigorous and costly RCTs with repeated assessments until it is better understood what aspects of UC may be most potent for improving secondary control coping.

CMs that did not significantly mediate treatment outcome were found in all three studies. In Study 1, family dysfunction reliably failed to mediate the effects of EBPs for youth depression in Study 1 by synthesizing data from 11 RCTs with over 1,000 depressed youths. As discussed earlier, this may not have been a surprising finding given the focus of EBPs for youth depression on the youth (vs. parents or family), thus its utility in informing future research on change processes and mechanisms may be minimal. However, the existence of a consistently nonsignificant CM serves to reinforce the specificity of the robust mediators—that EBPs are likely targeting certain change processes and mechanisms and not others. Other CMs that did not significantly mediate treatment outcome were identified from a smaller sample of RCTs and participants. Study 1 nonsignificant CMs based on meta-analysis of three to seven RCTs with fewer than 1,000 depressed youths include problem solving, pleasant activities, cognitive skills, and avoidant coping. Study 1 CMs that lacked sufficient data for meta-analysis, but were identified in the systematic review as having been tested as mediators by RCT authors, include role engagement, motivation to change, therapeutic alliance, expression of affect, group cohesiveness, and parent marital satisfaction. All were tested in one RCT. Study 2 and Study 3 nonsignificant CMs, also each assessed in a single RCT, were therapeutic alliance, parent- and youth-reported primary control coping, parent-reported secondary control coping, and parent-reported top problems. Using the optimal sample size for TSSEM as a rule of thumb, these variables tentatively failed to mediate the effects of EBPs for youth depression and thus require further research on the basis of sample size alone. Of course, there are other reasons that would
argue for continued study of these candidate mediators (CMs), such as possible differences between specific measures of each CM, or between EBPs, or between control conditions. Among these tentatively nonsignificant CMs, the ones that are predictors of outcome—pleasant activities, parent-reported primary control coping, and parent-reported secondary control coping—might be prioritized over those that are not in future research. Especially when the control condition is an active one such as UC, predictors may be potential common factors across different treatments. It is noteworthy that none of these nonsignificant CMs were significantly impacted by existing EBPs, but they might be by other treatments designed to target them (some of which could become EBPs in the future). Thus they could be considered inconclusive CMs of psychotherapy for youth depression more generally. True nonmediators that were changed by EBPs but nevertheless failed to predict outcome were not found in the three studies.

Having identified two robust mediators for further investigation as promising candidate change mechanisms, one consistently nonsignificant CM that should probably not be further pursued as a change mechanism of existing EBPs, as well as one tentative mediator and numerous tentatively nonsignificant CMs that require more research, I believe that I have achieved the first aim of my dissertation.

**Applying Advanced Quantitative Methods**

Given that I have applied different quantitative methods in the three studies to identify treatment mediators, I believe that I have also met the second aim of my dissertation at a basic level. However the different methods were used with varying success to address limitations of existing research.

In Study 1, the TSSEM approach to MASEM (Cheung, 2014; Cheung & Chan, 2005) was used quite effectively to expand the pool of RCTs and CM under consideration, with
missing data in matrices handled by maximum likelihood estimation, thereby facilitating the synthesis of data across multiple RCTs, and increasing power to identify the two robust mediators and one consistently nonsignificant CM. These robust mediators and consistently nonsignificant CM would not have been identified without MASEM. As useful as MASEM was for clarifying the mediator status of three CMs, it was of limited utility for many others that were measured in fewer RCTs, because the low sample size calls into question the reliability of the MASEM findings. The models for two of the CMs also did not fit the data well on some fit indices; presumably the models could be modified to improve their fit to the data. Finally, the low frequency of RCTs that assessed CMs during treatment precluded analyses assessing change in CMs before change in outcomes.

In Study 2, multiple imputation was first employed to maximize the use of available data in analyses while minimizing bias due to missing data (Raghunathan, Lepkowski, Van Hoewyk, & Solenberger, 2001; White, Royston, & Wood, 2011). Then mediation and moderated mediation were tested via bootstrapping, with 95% bias-corrected confidence intervals generated to account for the asymmetric distribution of the indirect effect, and thus increase power to detect mediators (Preacher & Hayes, 2008; Hayes, 2013). Here, it is unclear how successfully bias was minimized, but power was increased to the extent that the indirect effect and its confidence interval could be estimated even though the effect of treatment condition was nonsignificant. Mediation testing in the absence of a main effect, although not specific to bootstrapping, would have been ruled out in the popular causal steps proposed by Baron and Kenny (1986). Consequently, no mediators would have been identified from this study if Baron and Kenny’s method had been used. Similar to Study 1, the lack of mid-treatment measurements precluded analyses involving change in CMs before change in outcomes.
In Study 3, a whole slew of longitudinal growth models were fitted to take advantage of the weekly measurements of CM and outcome for clarifying temporal relationships between CM and outcome. Latent change score models (LCS; McArdle, 2009) were selected because they fit the parent-reported data better than other types of growth models and were more suited to evaluating the directionality of the CM–outcome relationship. Indeed, the LCS models revealed that the CM–outcome relationship was in the opposite direction from what was hypothesized. Unfortunately, the fit of the univariate LCS models was only marginally acceptable, and deteriorated in the bivariate LCS models, rendering the findings tentative. Moreover, none of the growth models fit the youth-reported data even marginally. The model fit issues may have stemmed from a high degree of between-youth heterogeneity in change trajectories and mean trajectories that are not well represented by standard change equations. Moreover, the relatively high proportion of missing data in some weeks may have exacerbated issues with model fit, even though missing data were ostensibly addressed with full information maximum likelihood estimation. Therefore, although Study 3 was proposed in order to demonstrate how temporal precedence of the CM might be assessed, its value may lie more in illustrating the substantial challenges that can arise from applying LCS and other longitudinal models to psychotherapy data. Publication bias virtually guarantees that the successful examples will be much more visible. The existence of such successful examples (e.g., Marker, Comer, Abramova, & Kendall, 2013) indicates that LCS models could help to answer important questions about how psychotherapies work with some datasets, but Study 3 shows that they may not be suitable for other datasets.

In sum, I think it would be fair to conclude that the quantitative methods were all well-equipped to examine mediation effects in the absence of the main effect of treatment
condition, and partially mitigated limitations relating to low power, small sample size, and missing data. Temporal relationships were not conclusively clarified due either to a lack of the requisite repeated mid-treatment measurements or to the nature of the repeated data.

A Research Agenda for Discovering How Psychotherapies Work

In planning a research agenda for the future, it may be helpful to refer to Doss’ (2004) research strategy based on his model of therapy change processes, client change processes, and change mechanisms. He proposed initially identifying effective treatments, then examining change mechanisms that precede those outcomes using mediation analysis, mapping pathways between client change processes and improvement in change mechanisms, and subsequently experimentally manipulating therapy change processes to assess their causal effects on client change processes. The field has made tremendous strides in the first step of identifying effective treatments through the EBP movement (e.g., Association for Behavioral and Cognitive Therapies & Society of Child Clinical and Adolescent Psychology, n.d.), but seems to have hit a wall in the second step of identifying mediators, especially those whose changes precede symptom improvement.

This dissertation suggested that parts of this methodological wall could be eroded through the quantitative methods, but that the temporal precedence criterion may be the highest part of the wall to scale in this second step. Fortunately, mobile technologies and intensive longitudinal data analytic approaches might provide the tools needed to accurately and feasibly measure and model change trajectories of CMs and outcomes. As described in the Study 2 and Study 3 discussion sections, ecological momentary assessment (EMA; Shiffman, Stone, & Hufford, 2008; see also Trull & Ebner-Priemer, 2013) could provide detailed data about the timing and shape of change of CMs and outcomes within and between individuals, and dynamic structural
equation models (DSEM) were developed precisely to analyze such intensive longitudinal data (Hamaker, Asparouhov, Brose, Schmiedek, & Muthén, 2017).

In addition, the National Institute of Mental Health (NIMH) is actively encouraging research corresponding to the third and fourth steps of Doss’ (2004) model. Their framework, termed experimental therapeutics (NIMH, 2016), emphasizes drawing on basic psychopathology research to measure change mechanisms at the biological level or through performance-based tasks, and may involve designing new interventions to target those change mechanisms. This framework builds on the NIMH’s Research Domain Criteria (RDoC) project, which champions research on domains of brain function (e.g., cognition, negative emotion) implicated in psychopathology, with a focus on how the underlying neurocircuitry relate to individual differences in treatment response, genetics, molecular, and social/environmental characteristics (Insel et al., 2010). Certainly, broadening the modalities of change process and mechanism measures beyond subjective report and observational ratings can only produce a more complete picture of how therapeutic gains are accomplished.

These new research endeavors form a critical component of the research agenda for discovering how psychotherapies work, and how to improve them; in my opinion, these efforts can be refined and enriched by continued efforts to harness the existing sources of data (Ng & Weisz, 2016, 2017; Weisz et al., 2013). My dissertation was only able to present a few different methodologies to more fully exploit collected data, but there no doubt exists other potentially useful methods. One such method is individual patient (or participant) data (IPD) meta-analysis (Cooper & Patall, 2009), which might allow more fine-grained mediation analysis if integrated with MASEM, especially the detection of moderated mediation, while also increasing sample size and the generalizability of findings across studies, treatments, and samples. Another method
is causal mediation analysis (Imai, Keele, & Tingley, 2010; Valeri & VanderWeele, 2013), which can identify mediators more likely to cause the outcome by quantifying the robustness to confounding of the mediator–outcome relationship.

A final item of my suggested research agenda involves considering and testing the possibility that individuals may differ in terms of how the same psychotherapy works for them. There has been a shift in all areas of healthcare from an evidence-based approach that emphasizes delivering the best-supported treatment on average to all individuals, to a personalized or precision approach that aims to select and deliver treatment in the way that would optimize therapeutic benefit for the individual (see Ng & Weisz, 2016, 2017). Developing and testing ways to personalize intervention effectively is an increasingly active area of research (e.g., Silberschatz, DeRubeis, McMain, & Levenson, 2016), and there are nascent efforts to examine candidate change processes and mechanisms that differ across subgroups of individuals (e.g., DeRubeis, Huibers, & Forand, 2016). Developing methods to elucidate how psychotherapies may work similarly or differently across subgroups of individuals, and applying these methods to prospective research or existing data, will be an exciting new frontier in psychotherapy research that could maximize treatment gains for many youths.
References


Hopkins University Press.


