Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations | Graduate School

2015

Therapist Adherence to Cognitive Behavioral Therapy for Anxious Youth Across a Case

Julia R. Cox
Virginia Commonwealth University

Follow this and additional works at: https://scholarscompass.vcu.edu/etd

 Part of the Child Psychology Commons, and the Clinical Psychology Commons

© The Author

Downloaded from
https://scholarscompass.vcu.edu/etd/3691

This Thesis is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

THERAPIST ADHERENCE TO COGNITIVE BEHAVIORAL THERAPY FOR ANXIOUS YOUTH ACROSS A CASE

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University

By: JULIA REVILLION COX
Bachelor of Science, Seattle University, 2009

Director: Michael Southam-Gerow, Ph.D.
Professor
Department of Psychology

Virginia Commonwealth University
Richmond, Virginia
April 2015

Acknowledgements

I would like to thank Drs. Michael Southam-Gerow, Bryce McLeod, and Elizabeth Farmer for their help developing this thesis and their invaluable feedback throughout the process, Adriana Rodríguez and Cassidy Arnold for their tireless coding efforts, and my family and friends for their support and generosity.

Table of Contents

Page

Acknowledgements ii

List of Tables iv

List of Figures. v

Abstract vi

Introduction 1
Why Measure Treatment Integrity? 3
Measurement of Treatment Integrity 8
Order of Therapeutic Practices 13
Coping Cat: Cognitive Behavioral Therapy for Child Anxiety 17

Method 21
Coping Cat Trial and Participants 21
Therapy Process Measures 23
Scoring Order 25
Analyses 32
Hypotheses 33

Results 34
Final Sample 34
Order Scores and Adherence 35
Order Scores and Client Characteristics 37
Item-Level Analysis 38

Discussion 40
Preliminary Findings 40
Limitations 42
Future Directions and Applications 44

List of References 48

Appendix 57

A Coping Cat Session Guide 57

Vita 58

List of Tables

Page

Table 1. Client Characteristics and Information from I-CBT Trial 22

Table 2. Demographic Information from Included and Excluded Samples 35

Table 3. Minimum, Maximum, and Mean Order Scores by Convention (n = 33) 36

Table 4. Correlation of Order Score Conventions (n = 33) 36

Table 5. Correlation of Child/Treatment Characteristics with Order Scores by Convention .. 37

Table 6. Maximum Values of CBAY-A Items Within Each Case (n = 33) 39

Table 7. Correlation of Maximum CBAY-A Item Values with Order Scores (n = 33) 40

List of Figures

Page

Figure 1. Content Only Scoring Convention 28

Figure 2. Skills Before Exposure Scoring Convention 29

Figure 3. Practice Items in Prescribed Order Scoring Convention 30

Figure 4. Practice Items in Prescribed Order, in Prescribed Time Scoring Convention 31

Abstract

THERAPIST ADHERENCE TO COGNITIVE BEHAVIORAL THERAPY FOR ANXIOUS YOUTH ACROSS A CASE

By Julia Revillion Cox, B.S.

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2015

Major Director: Michael Southam-Gerow, Ph.D.
Professor
Department of Psychology

The field has developed many evidence-based treatments (EBTs); the integrity of EBTs being delivered, however, has been studied less than rigorously. Because many treatment manuals are developed to be delivered session-by-session, one way to assess treatment adherence, specifically, is across the course of the case: do therapists deliver treatment components in the order prescribed? The goals of this study were to characterize how therapists deviate from prescribed order and how adherence to order relates to child characteristics. Therapy process data were collected from a subsample of children (N = 33, aged 7-15) that received cognitive behavioral therapy (CBT) to address primary symptoms of anxiety. Adherence to CBT was measured by the CBT Adherence Scale for Youth Anxiety (Southam-Gerow & McLeod, 2011). Four methods to assess order were developed. Analyses include descriptive and correlative

statistics that characterize the delivery of CBT and the relation between adherence to order and pretreatment characteristics.

Therapist Adherence to Cognitive Behavioral Therapy for Anxious Youth Across a Case

Research has established that approximately 20% of children and adolescents experience significant mental health problems (U.S. Public Health Service, 2000) that can result in individual distress and impairment (e.g., social difficulties, school failure, criminal activity, suicide; Mash & Barkley, 2003) along with enormous costs to society as a whole (e.g., criminal justice system, lost productivity, hospitalization, reliance on public assistance programs; World Health Organization, 2005). In the past few decades, the field has rushed to fill this void and address the complex and diverse clinical needs of this population with the development of therapies that target child and adolescent disorders. Indeed, the number of empirically-supported psychosocial treatment programs and packages expands each year—a recent review found over 350 treatments that have been tested in randomized controlled trials (Chorpita et al., 2011). More recently, however, as the field seeks to accumulate knowledge acquired from treatment outcome studies and transport those technologies to community settings, it has become clear that the rigor with which treatments are tested could be substantially improved.

One of the core tenets of experimental research is the importance of standardizing the independent variable across subjects: is the treatment being delivered to client A the same as what client B receives? Further, is what clients A and B actually received what they were *supposed* to receive (e.g., the treatment being tested)? The extent to which treatment outcome researchers can answer this question depends upon their measurement of treatment integrity, or treatment fidelity. While the exact terminology may not yet be consistent across the field, treatment integrity has come to be defined as "the extent to which the intervention was implemented as intended" (Perepletchikova & Kazdin, 2005). Recent scholarship has defined treatment integrity as including three main components: (1) *treatment adherence*, or the extent to

which prescribed strategies and content were delivered as intended; (2) *treatment differentiation*, or the extent to which the treatment being studied differs from a control intervention (i.e., treatment "purity"); and (3) therapist competence, how skillfully and responsively the therapist was able to deliver both the prescribed content and the relational components (Waltz, Addis, Koerner, & Jacobson, 1993; Perpletchikova & Kazdin, 2005; McLeod, Southam-Gerow, & Weisz, 2009). While treatment integrity is often considered an important component of initial efficacy trials, adequate measurement of treatment integrity in practice is largely lacking (Perepletchikova, Treat, & Kazdin, 2007). Further, measurement of and methods to promote treatment integrity have become hallmarks of dissemination and implementation science: as we continue to learn more about *how* to translate therapeutic innovations to the community more broadly, the question of *what* is actually delivered is increasingly salient.

The measurement of treatment integrity as it relates to experimental validity and the delivery of services more broadly has been less than rigorous. Currently in development, however, is a suite of treatment integrity measures that may provide a way to begin to address these concerns (see McLeod, Southam-Gerow, Tully, Rodriguez, & Smith, 2013). Using data collected with these measurement tools, the focus of this investigation is to describe the structure of treatment as it is delivered in an efficacy trial. The composition of modern treatment manuals is often strictly sequential, with each component or session building on previously covered material. Indeed, one must assume that the evidence-base for a given treatment is only supportive for the order described within the manual, but we cannot assume that the exact specification of the intervention is *required* to produce outcomes in service settings. So, the question remains: To what degree do therapists deviate from prescribed order and does it relate to any characteristics inherent to the child?

To this end, the present document will accomplish three goals: (1) Review the theoretical and definitional underpinnings of treatment integrity as it applies to experimental validity and to the translation of mental health technologies to community-based service systems; (2) Discuss the implications of maintaining—or eschewing—the prescribed order of treatment components; and (3) Propose a preliminary study of therapist adherence to order of treatment components with data from one randomized controlled trial.

Why Measure Treatment Integrity?

If a primary and basic goal of treatment outcome research is to describe and evaluate the efficacy and feasibility of a given protocol, it follows that treatment conditions should maintain internal validity. Strengthening the relation between the independent and dependent variables can be achieved by minimizing the possibility for alternative explanations of changes in the dependent variable (Kazdin, 2003). Pharmacological researchers, for example, must ensure that the chemical composition and potency of a drug being tested is carefully controlled and consistent across participants.

As with pharmacological research, the composition and potency of psychotherapy should remain constant across the cases within the treatment condition; psychotherapy, however, has proven difficult to measure as precisely as medication—and with good reason. Whereas medication may result in a predictable interaction of molecules within cells, given established knowledge of physical properties, psychotherapy is most often an interaction between people in a specific location within a specific period of time. Therapists' level of experience can vary widely (e.g., Brookman-Frazee, Garland, Taylor, & Zoffness, 2009; Addis & Krasnow, 2000), as can theoretical orientations (e.g., Baumann et al., 2006), and attitudes toward evidence-based practice (Addis & Krasnow, 2000; Aarons, 2004). Each of these factors can influence the

integrity of treatment as delivered. Clients, on the other hand, can present with complex psychopathology and severe functional challenges; further, clients may be unengaged and interpersonally difficult. The acceptability and *perceived* effectiveness of treatments likely play a large role in what treatment-as-delivered actually looks like. The structure and expectations of efficacy trials are such that unmotivated therapists and patients with inconvenient comorbidities are often excluded from the trial, presumably resulting in greater treatment integrity overall.

Despite the difficulties intrinsic to psychotherapy research, there are several methodological choices that scientists can make to maximize the internal validity of experimental treatment conditions within clinical trials. For example, careful and unbiased sampling of both study clinicians and participants can help ensure that there are no systematic differences between groups (Kazdin, 2003). Researchers can also insure that their clinicians receive standardized training and ongoing supervision (Bellg et al., 2004).

However, if a main goal of treatment outcome research is also to generalize and replicate the effects of a given treatment, external validity must also be considered (Kazdin, 2003). Should therapy procedures be too specific or samples too homogenous, the applicability of efficacy trials to children who access mental health services in the community can be severely limited. As Bellg and colleagues (2004) summarize, "Questionable internal and external validity may make it impossible to draw accurate conclusions about treatment efficacy or to replicate" (p. 444).

Perhaps the most widely used and easily implemented strategy to help manage both internal and external validity is the use of a highly operationalized and standardized treatment manual. Treatment manuals provide detailed session-by-session guides with prescribed—or, at the very least, suggested—content, activities, and strategies. Systematizing the independent variable in this way can also provide an enduring and disseminable template to facilitate

replication efforts. In short, standardized psychotherapy manuals codify *how* the treatment *should* be delivered, providing the 'treatment as intended' criterion against which therapist behavior can be compared.

The methodological importance of treatment integrity is clear: treatment integrity ensures that the treatment in question is the treatment being tested, reduces unintended variability, and can maximize effect sizes. But does a higher level of integrity translate to fewer symptoms and higher functioning in patients? It would seem the field has made an assumption: if an efficacious treatment has been identified, the treatment should be replicated accurately to achieve the full benefit and desired outcome. Contrary to this notion, the literature—albeit limited—is mixed on the direct relation between integrity and symptomatology: while several studies have strongly linked treatment integrity to therapeutic change (e.g., Erhardt, Barnett, Lentz, Stollar, & Raifan, 1996, Henggeler, Melton, Brondino, Scherer, & Hanley, 1997; Huey, Henggeler, Brondino, & Pickrel, 2000), others have found only limited supports (e.g., Sterling-Turner, Watson, & Moore, 2002; Jones, Wickstrom, & Friman, 1997), nonlinear trends (e.g., Hogue et al., 2008), or no support at all (e.g., Weisman et al., 2002; Bein et al., 2000) (see Perepletchikova & Kazdin, 2005 for a brief review).

It must first be noted that deviation from integrity cannot always be construed as a detriment to the 'quality' of treatment; indeed, departures from prescribed protocol to meet individual or population needs can augment treatment effectiveness. Flexibility is often encouraged by treatment developers (Kendall & Beidas, 2007) and has correlated with increased child engagement in later sessions (Chu & Kendall, 2009). Perhaps indicative of the benefits of therapist flexibility, Hogue and colleagues (2008) identified a curvilinear relationship between treatment adherence and internalizing outcomes among a sample of substance abusing

adolescents, such that lower and higher levels of adherence were associated with less therapeutic change compared to moderate levels of adherence, indicating that there may be an ideal middle ground.

Perepletchikova and Kazdin (2005) describe several variables that may contribute to variable treatment integrity-outcome relations. First, the characteristics of the treatment itself must be considered, including complexity (both for the therapist to deliver and for the client to receive), integration and use of resources, therapeutic modality (e.g., process oriented interpersonal vs. time-limited goal-directed cognitive behavioral therapy [CBT]), and acceptable timeline. For example, the integrity of structured treatments with a protocol that is easily followed (i.e., as in a comprehensive manual) is likely to be measured more easily because the criterion is so well defined. Process oriented and relational psychotherapies are much harder to specify, and therefore to measure with traditional experimental methodology (Perepletchikova et al., 2007).

Although its value in establishing experimental validity is well established, what usefulness might integrity measurement have in the delivery of mental health services more broadly? Whereas the goal of treatment outcome research is to develop etiological models and corresponding interventions to be tested and retested in an effort to explain and predict phenomena with precision, the explicit goal of the service system is to maximize client outcomes in an effort to increase quality of life (Regan, Daleiden, & Chorpita, 2013; Burns, Hoagwood, &Mrazek, 1999). In their recent manuscript, Regan and colleagues (2013) identify the common ground between these two goals: reducing or managing uncertainty. While it may not always be labeled as such, treatment integrity is a means to that end in both systems. For example, within the service system, administrators of community clinics attempt to maximize the odds of positive

client outcomes by making sure the organization provides the best possible services that are readily accessible and produce positive outcomes. Outcome and service utilization data are regularly collected to ensure the effectiveness of service provision. To this end, measurement of treatment integrity can be used to confirm that the treatments being delivered share characteristics with treatments that have been delivered successfully in other contexts (Regan et al., 2013).

In clinical research, the criterion (i.e., treatment protocol, therapist behaviors that are 'expected') is most often a foregone conclusion. Funding sources—along with the scientific method—require that treatment setting, method(s) of supervision, indicator(s) of psychosocial improvement, and measure(s) of outcome variables be carefully vetted and chosen—again, to help minimize uncertainty and optimize the effect of the independent variable. Stakeholders and leaders in the service system are faced with similarly complex and significant questions for a large number of diverse clientele with little guidance about how to best choose and implement effective treatments across settings. As McLeod et al. (2013) discuss, there is great potential for using treatment integrity measures to guide decision-making (i.e., feedback systems that utilize benchmarks), but as I will discuss shortly, measures of integrity with applicability in the service system are very limited.

Increased attention to and measurement of treatment integrity may also provide a yardstick with which to measure the value of treatments within the mental health care system more broadly. As the prevalence of 'pay for performance' and public accountability within the larger health care system increases (e.g., Rosenthal, Fernandopulle, Song, & Landon, 2004; Epstein, Lee, & Hammel, 2004), mental health care seems poised to follow. Widespread measurements of integrity could help consumers, policymakers, and other stakeholders make

more informed decisions about what providers and/or treatments to pursue and how much to pay for it. Relatedly, agency-led staffing decisions (i.e., hiring, promoting, and firing) may be increasingly influenced by integrity and outcome data. The lack of efficient and widely available measures of treatment integrity, however, seems to have stalled these developments somewhat (Schoenwald, 2011).

To continue exploring the value of treatment integrity as a means to maximize experimental validity and guide treatment implementation, I now turn to a discussion of treatment integrity as it has been measured.

Measurement of Treatment Integrity

How, given the complexity of psychotherapy, has treatment integrity been measured and described in the psychological literature most broadly? The short answer is not very well and not very consistently (Perepletchikova et al., 2007). Despite the methodological implications of integrity measurement, the field has been slow to adopt sophisticated measurement procedures. There are a number of reasons why this might be so (Perepletchikova & Kazdin, 2005): first—and perhaps most instrumental—is the lack of standardized, direct (i.e., unbiased observation) integrity measurement tools. The ability to rely on therapist or supervisor self-report is tenuous, at best, but the availability of reliable, well validated, and *practical* measures is scarce. Further, as I have mentioned above, the field has only recently begun to operationalize treatment integrity and come to a consensus about conceptualizing its components (e.g., Perepletchikova, 2011; McLeod et al. 2009).

Concerns about treatment integrity have been long standing, but early treatment outcome studies did not often employ treatment manuals nor was the nature of therapist training adequately described (Moncher & Prinz, 1991). In their review of treatment integrity checks,

Moncher and Prinz (1991) found that "the majority (55%) of the studies reviewed essentially ignored the issue of treatment fidelity" (p. 257) altogether. Treatment integrity monitoring in treatment outcome research was infrequent during the 1980s and early 1990s and the measures that did exist varied widely in format and focus, limiting the development of a cohesive definition of the construct (Waltz et al., 1993). For example, coders' level of experience and proximity to delivery of therapy varied widely (e.g., therapist-reported adherence vs. those who were experienced in delivering psychotherapy vs. those who were inexperienced delivering psychotherapy) similarly, the source of integrity information was not standardized—characterizing integrity from therapist process notes captures something very different from what can be observed by watching a videotaped session in its entirety (Waltz et al., 1993).

A more recent review of treatment integrity measurement in the literature has revealed that many of these inconsistencies persist (Perepletchikova et al., 2007). In their review of 147 randomized controlled trial papers from six high impact factor journals, it was revealed that only 3.5% of papers **established** (i.e., how researchers conceptualize integrity as well as the extent to which therapists were provided with a treatment manual and received training and supervision), **assessed** (i.e., use of direct and/or indirect observation, measurement of therapist adherence and/or competence, and the use of psychometrically valid and reliable integrity measures), **evaluated** (i.e., ensuring accuracy and representativeness of integrity assessment data, like rater independence, reliability, and training), and **reported** (i.e., the extent to which integrity procedures were reported) treatment integrity adequately. Among the papers selected, the *establishment* of treatment integrity was determined to be adequate with the highest frequency (15.8%), while the overwhelming majority of papers (87.6%) inadequately *evaluated* treatment integrity.

These data highlight several specific concerns about the state of treatment integrity measurement. For example, treatment integrity was not being measured in its entirety, with seemingly greater emphasis on adherence over competence and differentiation (Perepletchikova et al., 2007). Similarly, many papers reported assessment of treatment integrity using 'indirect' treatment- or trial-specific methods that are often limited to therapist, supervisor, and/or client self-report measures (rather than less biased and more preferable observational methods) and have limited psychometric support. As the authors (2007) observe, "measures of outcome receive far more attention than does treatment integrity" (p. 834); indeed, the "operational definitions and measures of reliability are detailed when behaviors serve as dependent variables and are virtually ignored when behaviors serve as independent variables" (p. 834), reflecting a "curious double standard" (Peterson, Homer, & Wanderlich, 1982).

While the low levels of adequate integrity measurement and reporting that are present in the clinical literature are certainly disappointing, there are a number of reasons why treatment integrity has been neglected. Building on the Perepletchikova et al. review (2007) of treatment integrity measurement, Perepletchikova, Hilt, Chereji, and Kazdin (2009) surveyed the corresponding authors of 90 RCTs for their perspectives on barriers to implementing treatment integrity assessment practices. Several 'strong' barriers were identified: these include (a) the lack of time and resources to collect adequate integrity data (i.e., across therapists, situations, cases, and sessions), (b) the lack of time and resources to design and validate integrity measures, and (c) the lack of time and resources to provide direct training to therapists. Concerns about study funds were joined by the popular belief that the construct of treatment integrity is unclear and that methods for integrity assessment have not been well enough defined (Perepletchikova et al., 2009).

In response to these findings, Perepletchikova and colleagues (2009) produced a set of recommended treatment integrity procedures for treatment outcome research. Suggestions include: (a) defining adherence, competence, and differentiation, (b) providing explicit description of treatment procedures, therapist training, and supervision, (c) direct assessment of treatment adherence and competence with reliable and valid measures across treatment phases, situations, sessions and/or cases, and (d) fully reporting on the integrity procedures used. While this list may be unduly optimistic given the resources required to fully implement such thorough integrity measurement procedures, Perepletchikova (2011) concedes that some procedures (e.g., use of a treatment manual, ongoing supervision and monitoring of treatment) are more essential than others (e.g., controlling measure reactivity). Even in settings where precision is paramount (i.e., efficacy trials), however, the field seems to be lacking (Perepletchikova et al., 2007).

While Perepletchikova and colleagues' (2007) review of the treatment literature included some community-based effectiveness trials, there are some unique challenges to integrity measurement that exist within community settings. First, community settings are often working with far fewer resources, including financial strain and a lack of trained personnel. The more time and resource intensive assessment strategies described by Perepletichikova et al. (2009; e.g., observational coding) may not be feasible in most service settings. Second, measurement strategies must be able to fit into the day-to-day operations of the agency as well as the established practice norms (Schoenwald et al., 2011). In addition to these constraints, assessment tools must also be psychometrically sound.

Ultimately, it is clear that integrity measures must be both efficient and effective (Schoenwald et al., 2011). This is a difficult task, and the field has been able to make only moderate strides toward these assessment tools. Schoenwald and colleagues (2011) highlight the

innovations of a few research teams. Sheidow, Donohue, Hill, Henggeler, and Ford (2008) were the first to develop and test a 'hybrid observational strategy' in which therapists and supervisors implemented an adherence monitoring system (i.e., in which therapists watch their own tapes of therapy and rate their own behavior) that was found to be reliable and feasible, if limited in its specificity. Supervision Relatedly, therapist self-report measures of practices (as opposed to adherence or nonadherence)—despite the clear bias pitfalls—may show some promise (e.g., the Therapy Process Checklist developed by Weersing, Weisz, & Donenberg, 2002; the Monthly Treatment and Progress Summary, developed by the Child and Adolescent Mental Health Division, 2003; the Consultation Record developed by Ward et al., 2013).

Some settings may be able to support observational assessment despite the resources it requires. The SafeCare ® parenting model (Lutzker & Bigelow, 2002), for example, has been implemented in a state child welfare system while promoting therapist adherence through live and/or recorded observation by a consultant or coach (Whitaker et al., 2012). It seems that tying integrity measurement to ongoing supervision or consultation provides a feedback loop that may facilitate greater therapist support (i.e., supervisors may be able to identify and reinforce adherent practices) while embedding treatment integrity measurement into the fabric of an agency (i.e., within supervision practices as usual).

Client report methods are also being explored. The Multisystemic Therapy (MST; Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 2009) team has developed its own parent-report adherence measure that relates to short- and long-term outcomes in youth (Schoenwald, 2008). The measure is used throughout the MST organization, but has little immediate applicability outside of the MST model.

Ultimately, integrity measurement in all settings is limited. Adherence and broader integrity measures are often protocol- or program-specific, reducing their utility. As the field continues to develop assessment tools to meet these needs, it is also important to consider the 'level' or scope of integrity (i.e., an individual session, across a case) that is most well suited to answer.

Order of Therapeutic Practices

In their recent paper, Regan et al. (2013) propose a framework through which the field can conceptualize treatment integrity as comparing what is observed (i.e., treatment as it occurs) to what is expected (i.e., what is described in a manual). This occurs on three different levels: (1) Within a single therapy session (i.e., one patient, one therapist, one clinic—the **"event-level"**); (2) Across therapy sessions (i.e., across a case, across therapists, across clinics; the **"episode-level"**); and (3) Across multiple types of therapy (the **"multi-episode-level"**). The development of this multi-level framework prompts the question: how has integrity been measured across each of these levels?

To begin to address this question, I built upon Perepletchikova and colleagues' (2007) review of treatment integrity measurement within the treatment outcome literature. Following the procedures outlined by the authors, I conducted a review of the more recent literature with the goal of identifying where, within Regan et al.'s (2013) framework, has measurement been conducted. Six psychology and psychiatry journals were identified as sources (*Archives of General Psychiatry*, *The American Journal of Psychiatry*, *The British Journal of Psychiatry*, *Journal of the American Academy of Child and Adolescent Psychiatry*, *Journal of Consulting and Clinical Psychology*, and *Journal of Clinical Psychiatry*; these were initially chosen by Perepletchikova et al. (2007) because of high impact factor and prevalence of treatment outcome

research papers). Picking up where Perepletchikova et al. (2007) left off, articles published in 2005 to 2012 were considered for inclusion.

The final sample of 112 papers were reports of randomized controlled trials that (a) assessed the effect of a psychosocial intervention on a set of dependent measures (i.e., outcome measures), (b) included a comparison control group, (c) research was prospective and included randomized assignment of participants, (d) participants were recruited because they were experiencing psychological problem(s) of some sort, and (e) post-treatment outcome measures were assessed. Papers were excluded if (a) their primary purpose was anything other than evaluating the effect of a psychosocial treatment on outcome measures (e.g., identification of moderators, mediators), (b) interventions were not delivered by therapists (e.g., bibliotherapy, computerized therapy), (c) the comparison was between a highly standardized and a low standardized treatment, and (d) if only pharmacological interventions were evaluated.

Given these criteria, 112 treatment outcome research studies were conducted; 35 of these (31.3%) were child therapy-focused. Once identified, papers were coded for the population of focus (i.e., adult or child/adolescent), the target problem, the extent of integrity measurement across Perepletchikova et al.'s criteria (i.e., establishing, evaluating, assessing, and reporting treatment integrity), and the level at which integrity was measured (i.e., session-by-session or across the course of a case).

Of the identified papers, only 47 (42.0%) described how therapists were trained, and 50 (44.6%) described how therapists were supervised (i.e., how treatment integrity was established). Only 24 (21.4%) studies assessed treatment adherence using validated measures while only 25 (22.3%) studies assessed treatment adherence using reliable measures. Further, only nine studies (8.0%) reported on treatment adherence in a way that was informative of adherence levels (i.e.,

not simply ‘present’ or ‘absent’). Forty-eight studies (42.9%) did not monitor integrity measurement at all. Of the 64 studies that measured treatment integrity in some way, only one (1.6%) measured integrity across a case and 61 (95.3%) measured integrity session-by-individual-session, indicating that most formal integrity measurement to date has been, as defined by Regan et al. (2013), at the event-level: is the therapist doing what he or she should be doing *in this session*, rather than at the episode-level, or within the greater context of the case?

Despite the reliance on event-level integrity measurement, the expansion of the integrity framework to include episode- and multi-episode-levels allows for researchers and stakeholders to begin asking questions about the structure of treatment and the coordination of treatment components. For example, were all components delivered to the client in sequence? And, perhaps most importantly, does sequence matter?

The extent to which these questions are useful depends upon the rationale behind the prescribed sequencing of events within a given treatment protocol. Modern treatment manuals are most often linear session-by-session guides of content. It may be that there is a theoretical reason for placing one component before another. For example, in some cognitive behavioral therapy (CBT) for child anxiety programs, therapists teach cognitive coping skills *before* starting exposure sessions in an effort to increase a child’s sense of self-efficacy and increase the likelihood that he or she will experience early successes (e.g., Kendall, 1994). Likewise, many behavioral parent training programs prescribe that therapists first teach parents to facilitate the development of the parent-child relationship through child-directed positive play and praise *before* discipline training, in an effort to provide a greater incentive to the child for compliant behavior and increase the likelihood of success (e.g., Parent-Child Interaction Therapy, Eyberg, Boggs, & Algina, 1995; Triple P: The Positive Parenting Program, Sanders, 1999). In some

programs, however, the theoretical reasoning may not be clear—or may be totally absent (e.g., organizing skills training around an acronym).

The development of "modular" treatments (e.g., the Modular Approach to Therapy for Children with Anxiety, Depression, Trauma, or Conduct Problems [MATCH]; Chorpita & Weisz, 2009), in which the sequence of treatment components is organized around an adaptive algorithm, allow for increased therapist flexibility based on feedback on outcomes. For example, a therapist treating an adolescent presenting with primary depression may be able to integrate anxiety-focused treatment components if the adolescent also experiences anxiety that interferes with depression-focused treatment. Indeed, a recent effectiveness trial found that a modular protocol outperformed linearly organized "standard" EBTs and usual care in community- and school-based mental health settings (Weisz et al., 2012). The provision of flow charts and algorithms provides a more complex comparator of prescribed sequencing, allowing for several permutations of treatment that can be considered adherent.

Measurement of adherence at the episode-level is a subject of increasing interest in the field. In one of the first empirical investigations of practice sequencing, Park and colleagues (2015) catalogue the frequency of adherence to practice sequencing in both modular and linear treatments. The authors created scoring rules based upon the modular algorithms and the standard treatment manuals. Using the therapist-completed Consultation Record (Ward et al., 2012) as a present/absent measure of treatment integrity (i.e., the therapist reported that he/she completed most of the main steps associated with the treatment component), Park and colleagues (2015) identified the number of sessions that adhered to the expected components in a given session based upon what components were present in the session immediately preceding. They determined that 43.6% of 924 standard treatment sessions and 32.3% of the 873 modular

treatment sessions deviated from the standard linear progression and the modular algorithm, respectively. These results suggest that although therapists delivered practice content, but deviance from prescribed practice arrangement was a relatively common practice across both conditions of the trial. Therapists commonly skipped ahead or behind in the treatment protocol, integrated other non-prescribed practices, and covered practices from a different problem area (e.g., delivering an anxiety-focused module within a case focused on conduct problems).

As discussed above and illustrated by the work of Park et al. (2015) and others, psychotherapy is not a manufacturing process: therapists are encouraged to be flexible with their clients, adapting the treatment to meet the specific needs of the youths they see. How *extensively*, then, do therapists deviate from the prescribed order?

The present study seeks to begin to address this question. Within the Regan et al. (2013) framework, it is clear that these questions are best answered at the episode-level, or across the multiple sessions that comprise a case. Because the variable of interest is discrepancy between what is observed and what is prescribed, this study will also require a measure of treatment adherence (i.e., did the therapist deliver the treatment as intended). While a close examination of treatment sequence may benefit the understanding of any treatment literature, CBT for child anxiety is an ideal place to start.

Coping Cat: Cognitive Behavioral Therapy for Child Anxiety

Given the importance of exploring the construct of integrity as it relates to the order of therapeutic strategies, it would seem that the body of research in CBT for child and adolescent anxiety is particularly well suited for an initial investigation. First, there is the potential for substantial real world impact, as anxiety disorders are among the most prevalent psychiatric disorders experienced by children and adolescents. Anxiety disorders (i.e., Generalized Anxiety

Disorder [GAD], Separation Anxiety Disorder [SAD], Social Phobia [SP]) occur in between 6% and 15% of the child and adolescent population (Bernstein & Borschardt, 1991; Anderson, Williams, McGee, & Silva, 1987; Fergusson, Horwood, & Lynskey, 1993). Youths with these problems can also experience substantial impairment in day-to-day school, family, and interpersonal functioning (e.g., McClure & Pine, 2006; Silverman & Berman, 2001).

Building on cognitive-behavioral theory and social learning principles, the Coping Cat program was the first manual-based treatment developed and tested for anxiety disorders in youth, ages 7 to 13 (Kendall, 1994). It has been tested extensively; evidence for its efficacy includes three randomized controlled trials of its individual format (Kendall, 1994; Kendall et al., 1997; Kendall, Hudson, Gosch, Flannery-Schroeder, & Suveg, 2008), two trials of Coping Cat delivered in a group format (Flannery-Schroeder & Kendall, 2000; Kendall et al., 2008), and a pilot study of a modified, emotion-focused Coping Cat protocol (Suveg, Kendall, Comer, & Robin, 2006). Children and adolescents included in these trials presented with several permutations of anxiety, including GAD, SAD, SP, and earlier diagnostic analogues (e.g., Overanxious Disorder, Avoidant Disorder from the Diagnostic & Statistical Manual of Mental Disorders, 3rd ed.; American Psychiatric Association, 1980) and the majority of participants (64% - 73%) experienced substantial symptom reduction post treatment (Kendall, 1994 and Kendall et al., 1997, respectively) and in one trial (Kendall, 1994), many of these gains were maintained between two and five years post (Kendall & Southam-Gerow, 1996).

The standard protocol tested in these trials has a prescribed order of content that is designed to require approximately 16 treatment sessions to complete (Kendall, 1990; Kendall & Hedtke, 2006). The first phase of treatment is designed to help children build cognitive and behavioral skills that will help them cope with anxiety. This multi-step coping plan is known as

the FEAR Plan, in which clients are taught: (a) how to identify the physiological sensations associated with their anxiety (**F**eeling frightened?) and to address these feelings with relaxation exercises; (b) how to identify cognitive distortions (e.g., maladaptive expectations; **E**xpecting bad things to happen?); (c) how to challenge those cognitions and use problem solving to find the best possible coping strategy (**A**ttitudes and **A**ctions that might help.); and, finally, (d) how to monitor their progress and reward their efforts to face their fears (**R**esults and **R**ewards), with the intention of building self-esteem and -efficacy and shaping behavior away from avoidance (Kendall, Furr, & Podell, 2010).

Children then begin the exposure and practice phase of treatment, during which the therapist helps the child identify a fear hierarchy to help guide systematic desensitization to the feared stimuli via imaginal and in vivo exposure tasks. Using the FEAR Plan to problem solve and restructure cognitions along the way, the child moves from feared situation to increasingly feared situation, with the goal of habituation and, ultimately, mastery of feared tasks/situations (Kendall et al., 2010). The rationale behind the order of these components is—at a more global level—to first build rapport and self-efficacy before beginning the much more taxing exposure tasks.

In the decades since the first efficacy trial, Kendall and colleagues have explored the opportunities for flexibility within the protocol—and, indeed, 'flexibility within fidelity' is explicitly encouraged (e.g., Kendall, Gosch, Furr, & Sood, 2008; Chu & Kendall, 2009). Acceptable flexibility can include: the modification of treatment content/complexity to be (a) more developmentally appropriate (i.e., younger children or children with cognitive limitations may require a more concrete and behaviorally-focused approach); and (b) more accessible to children with skills deficits or comorbidities (i.e., children may need additional emotion

education or practice in social situations; children with comorbid ADHD may need more frequent breaks and shorter sessions) (Beidas, Benjamin, Puleo, Edmunds, & Kendall, 2010). Flexible applications of Coping Cat can remain adherent if (1) the stated session goals are being met (i.e., skills-related content was delivered, exposure tasks were completed); (2) the child's anxiety and the application of any modifications are conceptualized within the cognitive-behavioral perspective; (3) treatment is active, moving toward the goal of desensitization toward feared stimuli (i.e., children engage in exposure tasks and the avoidance or escape of feared stimuli is not reinforced); and (4) clinicians adhere to social learning theory (i.e., using praise to shape behavior, assigning homework to help the child practice generalize freshly learned skills) (Kendall et al., 2008).

Despite the built in flexibility and empirically supported promise of CBT for child anxiety programs, when delivered by clinicians in community mental health clinics, it has not outperformed usual care (Southam-Gerow et al., 2010; Barrington, Prior, Richardson, & Allen, 2005). Following their rigorous examination of Coping Cat vs. treatment as usual in community clinics, Southam-Gerow and colleagues (2010) discuss several reasons why the implementation of CBT may have influenced these results. Perhaps the most salient, given our discussion about flexibility within fidelity, is that CBT therapists maintained a focus on anxiety whereas usual care therapists were able to flexibly respond to any comorbidities or crises that arose. Additionally, CBT therapists had variable experience delivering CBT while usual care therapists were able to use strategies about which they felt comfortable and confident.

Coping Cat, as a well-tested and well-described protocol, is an excellent place to start. Coincidentally, session data from two Coping Cat trials are in the process of being coded for the

development of treatment integrity measures, fulfilling this study's need to assess adherence across a case.

Method

Coping Cat Trial and Participants

As a part of the Treatment Integrity Measurement Study (TIMS), a larger NIH-funded integrity measurement development study (NIMH RO1 MH086529; see McLeod et al., 2013), therapy process data were collected from a total of 89 youths, age 7 – 15, who participated in one of two randomized controlled trials of the Coping Cat program for the treatment of anxiety disordered children and adolescents. First, the Coping Cat Study (Kendall, Hudson, Gosch, Flannery-Schroeder, & Suveg 2008) compared the efficacy of individual-CBT, family-CBT, and an active control (i.e., family-based education, support, and attention). Children and families were treated in an outpatient facility at the Child and Adolescent Anxiety Disorders Clinic at Temple University (see Kendall et al., 2008 for details). Second, the Youth Anxiety Study (YAS; Southam-Gerow et al., 2010) compared the effectiveness of individual-CBT and usual care. Children were treated at one of six community-based outpatient clinics in Los Angeles County (see Southam-Gerow et al., 2010 for details).

The present study focused only on the children and adolescents participating in the individual-CBT condition of the Coping Cat Study (I-CBT; $n = 55$). Among these youths, participants were, on average, 10.6 years old (SD = 2.0). All participants in the sample met diagnostic criteria for a primary anxiety disorder. Please see Table 1 for additional participant information.

Table 1.

Client Characteristics and Information from I-CBT Trial

Variable	CC (*n* = 55)
Youth age in years *(SD)*	10.4 *(1.9)*
Percent of female youths	39.2
Percent of Caucasian youths	86.3
Percent of youths with principle diagnoses	
GAD	47.1
SAD	41.2
SP	-
SOP	45.1
Diagnoses at initial assessment *(SD)*	3.56 *(1.74)*
Percent of family income	
Up to $60,000/year	35.3
Above $60,000/year	56.9

Note. GAD = generalized anxiety disorder, SAD = separation anxiety disorder, SP = specific phobia, SOP = social phobia. A number of youths had more than one primary diagnosis.

As described above, these youth received individual child-focused Coping Cat, a manualized CBT program designed to treat youth with anxiety disorders (Kendall & Hedtke, 2006; Kendall, 1994; Kendall, Kane, Howard, & Siqueland, 1990). While the Coping Cat program is intended to be delivered in 16 – 20 sessions, CC participants received 16.7 sessions on average (SD = 1.43; range 8 – 18) over the course of 19.5 average weeks (SD = 3.9; range 8 – 30). Of the youth who began treatment in the YAS trial, participants received 16.8 sessions average (SD = 5.0; range 2 – 22) over the course of 26.21 average weeks (SD = 12.6; range 2 – 48). At the end of treatment, 64% of youth no longer met criteria for their primary anxiety diagnosis post treatment.

Above and beyond child demographic data, Kendall et al. (2008) also administered the Child Behavior Checklist (CBCL), a 118-item parent-reported measure of clinically relevant child behavior. The CBCL is widely used in both clinical and research settings and has amassed a great deal of evidence supporting its validity and reliability (Achenbach, 1991). Many psychometrically sound and clinically useful (e.g., Connor-Smith & Compas, 2003) subscales can be gleaned from this measure, including two broadband factors and eight syndrome scales that are age- and gender-normed. The mother-reported Internalizing Problems scale was used to assess anxiety-related symptomatology pre-treatment.

Participating children and adolescents were seen by a total of 16 therapists. Therapists were masters-level clinicians enrolled in the clinical psychology doctoral program at Temple University who, upon completing approximately six hours of training in Coping Cat, received ongoing weekly supervision with experienced doctoral-level psychologists (Kendall et al., 2008).

Therapy Process Measures

For the purposes of facilitating clinical supervision and/or monitoring treatment adherence (i.e., as an independent variable check), all sessions were audio or video taped for the Coping Cat trial. Of the original 812 I-CBT sessions, 532 (65.5%) sessions were coded as a part of the larger TIMS project. Sessions were excluded for several reasons, including: fewer than two audible sessions, unintelligible recordings, and receipt of treatment from more than one therapist.

Sessions were coded using four observational measures that correspond to the three dimensions of treatment integrity, namely: (1) therapist **adherence** to cognitive-behavioral treatment components for anxiety with the CBT Adherence Scale for Youth Anxiety (CBAY-A; Southam-Gerow & McLeod, 2011; Southam-Gerow et al., in press); (2) therapist **competence**

delivering cognitive-behavioral components for anxiety with the CBT for Youth Anxiety Competence Scale (Southam-Gerow & McLeod, 2012) and delivering 'common factors' with the Common Factors Therapist Competence Scale (Brown, Southam-Gerow, & McLeod, 2012); and (3) therapist delivery of additional treatment strategies with the Therapy Process Observational Coding System for Child Psychotherapy-Strategies Scale (McLeod & Weisz, 2010), a measure of **differentiation**. Given the importance of measuring order or arrangement of CBT-specific strategies, the CBAY-A is the focus of the current study.

CBT Adherence Scale for Youth Anxiety. The goal of the CBT Adherence Scale for Youth Anxiety (CBAY-A) is to "provide comprehensive descriptive data on the breadth and depth of CBT interventions delivered to children and families" (Southam-Gerow & McLeod, 2011, p. 3). The measure was designed to capture the *content* of therapist behavior—as opposed to how *well* or *competently* specific strategies were delivered—regardless of client behavior (e.g., enthusiastic participation) and/or outcomes. In this way, it provides a session-by-session snapshot of CBT content dosage.

The CBAY-A is a 22-item observational measure comprised of three code types: (1) Standard Items, therapeutic strategies that are expected to occur in most or all CBT for child anxiety sessions (e.g., setting a session agenda, checking and assigning homework); (2) Model Items, therapeutic strategies that are specific to CBT for child anxiety (e.g., psychoeducation, building an exposure hierarchy, teaching a multistep coping plan); and (3) Delivery Items, identifying *how* therapists teach Model Items (e.g., didactic teaching strategies, asking a child to rehearse behavioral skills). Items each receive a score on a 7-point 'extensiveness' scale, such that items were delivered 1 = *not at all*, 3 = *somewhat*, 5 = *considerably*, and 7 = *extensively*. 'Extensiveness' is comprised of two separate dimensions observed across the course of an

individual session: (1) thoroughness, as established by the effort or commitment the therapists put toward the intervention strategy; (2) the detail with which the therapist describes the reasoning for the intervention; (3) the depth/intensity of the intervention; (4) the degree to which the therapist 'follows-through' with the intervention; and (5) how persistently the therapist pursues the intervention across the session (Southam-Gerow & McLeod, 2011). Preliminary reliability and validity data were supportive for item scores, indicating that the CBAY-A is a promising observational measure of therapist adherence to individual CBT for child anxiety (Southam-Gerow et al., in press).

Coding procedures. Two doctoral students in clinical psychology independently coded all eligible I-CBT (n = 532) sessions. The principal investigators of the TIMS project trained both coders over the course of six months to reach adequate pre-study reliability at the item level, such that ICC(2, 2) > .59 (indicating "good" or "excellent" agreement; Cicchetti, 1994). Training involved: (1) didactic instruction and discussion regarding CBT for child anxiety and the Coping Cat manual (Kendall, 1994) paired with joint coding exercises to explore each item; (2) independent practice coding, and (3) establishing reliability with each other and with master coders on 32 recordings. Once evidence for good reliability was demonstrated under these conditions, again ICC(2,2) >.59, coding began in earnest. For the duration of the coding period, coders met with the principal investigators every two weeks to prevent coder drift (Margolin et al., 1998). Reliability coefficients for newly coded tapes were calculated for each meeting and, if ICCs fell below the threshold or had declined consistently, additional training was provided.

Scoring Order

Because adherence is most often conceptualized at the event-level, there are no well-established methods for measuring the order of therapeutic operations across the course of a case

at the episode-level. Further, measures of adherence have most often been binary, capturing only if the content was delivered or not, rather than a more nuanced illustration of adherence on a continuum. The extensiveness scores derived from the CBAY-A are designed to capture these distinctions, but require a more complicated scoring procedure to glean meaningful information about the relative level of adherence to the manual at any given point in treatment. In an attempt to capture both content and sequencing of practice elements as measured by the Model Items, here I present four different methods for measuring adherence to arrangement of practices that vary by degree of flexibility. In an effort to streamline these procedures to most efficiently answer the questions at hand, three modifications were made: (1) the first (Session 1) and last (Session 16) sessions were excluded due to a large amount of missing data, (2) parent-only sessions (Session 4 and Session 8) were excluded, leaving the content of 12 child-only sessions to be scored, and (2) a composite Exposure item was created by averaging three exposure task items, namely Exposure Preparation, Exposure, and Exposure Debrief items. Preliminary analyses determined that these three items were significantly correlated, ranging from $r(742) = .69$ to $r(742) = .81$ ($p < .001$), and a combined score could more effectively capture exposure tasks that happened outside of the therapy room and were therefore not coded (e.g., preparing the child for a social anxiety exposure, leaving the room to complete the task, and returning to the room to discuss and debrief).

To guide the scoring process, an expert in the Coping Cat program identified the top two CBAY-A items that one would expect to see in each Coping Cat session (see the Session Guide, Appendix A). For example, during Session 7, therapists are expected to teach their clients a multi-step problem-solving plan, a strategy that is part of the "A" step in the larger FEAR coping plan. Given this session content, I expect to see the "Problem-solving" and "Coping plan" items

coded. The extensiveness score given to each of these identified items translates into a point value (i.e., extensiveness score of "1" results in one point, "2" results in two points, etc.). Therefore, there are a total of 14 points to be earned during each session. Each case will be given a point value (24 to 168 points possible). To place this score on a common and more meaningful scale (namely, the 1 to 7 extensiveness score employed by the CBAY-A measure itself), the total number of points will be divided by 24 (i.e., the total number of items being scored), producing an average extensiveness score for adherence to order, with "1" indicating the therapist never delivered the correct intervention and "7" indicating that the therapist extensively delivered every intervention we expect to see.

Below each of the four scoring conventions are described in greater detail. All will be applied to each case, such that each case will have four corresponding order scores that reflect scoring conventions of increasingly reliance on the order prescribed by the manual.

Content only. The most liberal of these scoring methods is concerned only with content: did the therapist deliver what he or she was supposed to deliver? At this scoring level, the threshold is simply delivering the material at some point during the case. Following the expected session content outlined by the Session Guide (Figure 1), each CBAY-A model item was identified as expected in a set number of sessions (e.g., the Psychoeducation item is expected in one sessions, while the Exposure Composite item is expected in six sessions). Because expected order is irrelevant at this level, the highest extensiveness scores will be selected for each item. For example, Fear Ladder is expected to appear in one sessions; if, over the course of the case, Fear Ladder has been coded in a total of four sessions, the highest extensiveness score will be used to calculate points. See Figure 1 for a summary of the scoring convention at this level.

Content Only Scoring Convention		
Item	**Sessions Present**	**Points Possible**
Psychoeducation	1	7
Emotion Education	3	21
Fear Ladder	1	7
Relaxation	1	7
Cognitive-Anxiety	1	7
Problem Solving	1	7
Self-Reward	1	7
Coping Plan	9	63
Exposure Composite	6	42
Total:	24 scores	168
Divide number of points by 24 scores to get extensiveness score on 1-7 scale.		

Figure 1. Content Only Scoring Convention

Skills before exposure. This scoring method evaluates order based on treatment structure that was largely guided by theory: did the therapist deliver skills training before starting exposure? To receive points at this level, only skills training points obtained before exposure tasks began will be counted toward the overall score. All Psychoeducation, Emotion Education, Fear Ladder, Relaxation, Cognitive-Anxiety, Problem Solving, and Self-Reward points (77 total) along with 21 Coping Plan points are available to be earned, but only before Exposure is scored. For example, if the therapist begins exposure before delivering relaxation, but returns to relaxation training later in the case, no Relaxation item points will be awarded. After Exposure has been coded, only Exposure and seven additional Coping Plan points can be awarded. See Figure 2 for a summary of scoring at this level.

Skills Before Exposure Scoring Convention		
Item	**Sessions Present**	**Points Possible**
Psychoeducation	1	7
Emotion Education	3	21
Fear Ladder	1	7
Relaxation	1	7
Cognitive-Anxiety	1	7
Problem Solving	1	7
Self-Reward	1	7
Coping Plan	3	21
Pre-Exposure Total (84) must be earned before Exposure Coded		
Coping Plan	6	42
Exposure Composite	6	42
Total:	24 scores	168
Divide number of points by 24 scores to get extensiveness score on 1-7 scale.		

Figure 2. Skills Before Exposure Scoring Convention

Practice items in prescribed order. This scoring method requires that content be delivered in the order they were written in the manual, regardless of session number: did the therapist deliver therapy content in order? To receive points at this level, only the extensiveness ratings obtained before the next unique Model Item receives points. For example, if Relaxation is coded before Cognitive-Anxiety appears, points will be awarded. However, if Cognitive-Anxiety is delivered after Problem Solving has been coded, no points will be given for Cognitive Anxiety. Although this scoring convention is substantially stricter than the first two described here, therapists are not required to complete session content within a prescribed number of sessions (e.g., spending two sessions on relaxation instead of one). See Figure 3 for a summary of scoring at this level.

Practice Items in Prescribed Order Scoring Convention		
Item	**Sessions Present**	**Points Possible**
Psychoeducation	1	7
Emotion Education	2	14
Fear Ladder	1	7
Pre-Relaxation Total (28) must be earned before Relaxation Coded		
Relaxation	1	7
Emotion Education	1	7
Pre-Cognitive Total (42) must be earned before Cognitive Coded		
Cognitive-Anxiety	1	7
Coping Plan	1	7
Pre-Problem Solving Total (56) must be earned before Problem Solving Coded		
Problem Solving	1	7
Coping Plan	1	7
Pre-Self-Reward Total (70) must be earned before Self-Reward Coded		
Self-Reward	1	7
Coping Plan	1	7
Pre-Exposure Total (84) must be earned before Exposure Coded		
Coping Plan	6	42
Exposure Composite	6	42
Total:	24 scores	168
Divide number of points by 24 scores to get extensiveness score on 1-7 scale.		

Figure 3. Practice Items in Prescribed Order Scoring Convention

Practice items in prescribed order and in prescribed time. Scoring at this level requires that content be delivered in order and within the expected number of sessions: did the therapist deliver content within the allotted time? To receive all possible points at this level, the therapist will have had to follow the Coping Cat manual very closely and all possible points must be earned within 12 total sessions. For example, the 7 possible Psychoeducation points can only be earned during Session 3; further, the 63 possible Coping Plan points can only be earned during Sessions 6, 7, 8, and 10 through 15. See Figure 4 for a summary of scoring at this level.

Practice Items in Prescribed Order Scoring Convention		
Item	**Session #**	**Points Possible**
Emotion Education	2	7
Fear Ladder		7
Session 2 Total (14) must be earned before Session 3		
Psychoeducation	3	7
Emotion Education		7
Session 2-3 Total (28) must be earned before Session 5		
Relaxation	5	7
Emotion Education		7
Session 2-5 Total (42) must be earned before Session 6		
Cognitive-Anxiety	6	7
Coping Plan		7
Session 2-6 Total (56) must be earned before Session 7		
Problem Solving	7	7
Coping Plan		7
Session 2-7 (70) must be earned before Session 8		
Self-Reward	8	7
Coping Plan		7
Session 2-8 (84) must be earned before Session 10		
Coping Plan	10	7
Exposure Composite		7
Session 2-10 (98) must be earned before Session 11		
Coping Plan	11	7
Exposure Composite		7
Session 2-11 (112) must be earned before Session 12		
Coping Plan	12	7
Exposure Composite		7
Session 2-12 (126) must be earned before Session 13		
Coping Plan	13	7
Exposure Composite		7
Session 2-13 (140) must be earned before Session 14		
Coping Plan	14	7
Exposure Composite		7
Session 2-14 (154) must be earned before Session 15		
Coping Plan	15	7
Exposure Composite		7
Total:	24 scores	168
Divide number of points by 24 scores to get extensiveness score on 1-7 scale.		

Figure 4. Practice Items in Prescribed Order, in Prescribed Time Scoring Convention

Analyses

Data considerations and preparation. Because the data used in the present study are based upon archival data culled from an independent randomized controlled trial, there are idiosyncrasies to address.

Missing data. First, I accounted for missing data. As described above, a number of sessions that occurred were not available to be coded (280 [34.49%] I-CBT sessions total). Because I am interested in what happens over the course of a case, it is important to ensure that the order scores calculated capture the content and order of therapeutic interventions accurately.

I first minimized the impact of missing data by only including cases for which 75% of the 12 scored sessions were coded (n = 33 cases). Among the final sample, however, I was forced to account for session-level data that does not exist to create comparable order scores across cases. Unfortunately, because order or arrangement of prescribed practices remains understudied, there is little guidance within the literature to guide missing data practices. To account for and capture systematic differences that may occur at the therapeutic strategy-level (e.g., higher extensiveness scores of the Exposure Composite item is not expected to appear until the latter half of treatment) and at the item-level (e.g., exposure may be more challenging to implement in session than skill building components), I gleaned information from the cases for which there were complete data ("complete cases"; n = 4). Using a modified 'hot-deck' imputation approach (Allison, 2002) I imputed extensiveness scores from the complete cases from each session that was missing. For example, if Session 3 were missing from Case A, each coded item from the complete case set were averaged and imputed. If Sessions 5 and 10 were missing from Case B, each coded item from Session 5 of the complete case set were averaged and imputed for that session while each coded item from Session 10 of the complete case set were averaged and imputed for that session.

I chose this imputation strategy because it addresses item-level differences and allowed me to pull nuanced information from more sessions (i.e., not just the previous session or collapsing all items across the case) to capture what I hope approximates the mean dose of therapeutic strategies that the child received during a missing session.

Primary analyses. The primary aim of this study was to propose a new way to consider and measure treatment integrity, namely within the broader scope over the course of a case. After accounting for missing data, I used these data to describe *how* therapists adhere to order. Then I ascertained how adherence to prescribed order relates to child characteristics. These analyses are described in greater detail below.

Descriptive analyses were employed to determine if and how therapists adhered to or deviated from order. For example, were skills building sessions often delivered out of the prescribed order? Did therapists cover all skills with equal extensiveness? Pulling data from an efficacy trial allowed for a snapshot of manualized therapy in efficacy trials. If the I-CBT trial may be considered the 'ideal' circumstance for delivery of CBT for child anxiety, how frequent and extensive were deviations from the prescribed order? We also explored client and treatment variables that may help explain how and in what order therapists delivered CBT. To meet these goals, we present descriptive analyses at different levels, including mean scores (e.g., at the order score-level, the CBAY-A item-level). Further, we present correlative data to identify relations between order score and child-level demographic information.

Hypotheses

Because order has remained a largely ignored component of treatment integrity, there are, to our knowledge, no well-established methods of ensuring measuring adherence to and deviance from the prescribed order of treatment strategies using a scale such as the CBAY-A. I was,

therefore, without much guidance about how to best develop such a measurement system. Given the exploratory and largely descriptive nature of this study, formal hypotheses about the predictive utility of order are premature.

Results

Final Sample

As described above, the sample included in these analyses (n = 33) was culled from the larger intervention group (n = 55) due to the availability of session-level data. These children were seen by a total of 14 therapists, each of whom saw between one and six (M = 2.00, SD = 1.00) children included in this sample. The subsample chosen for this project accounted for 60.00% of the total trial sample. To investigate the possibility of any systematic differences between those included and excluded from this project, I compared the groups on a number of demographic and treatment characteristics. Independent *t*-tests and Chi-square analyses revealed no significant differences at the p = 0.05 level. Demographic information for the included and excluded subsamples are presented in Table 2 below.

Table 2.

Demographic Information from Included and Excluded Samples

	Included Sample (n = 33)		Excluded Sample (n = 22)	
Variable	M (SD)	n (%)	M (SD)	n (%)
Age	10.20 (1.59)	-	10.64 (2.25)	-
Female	-	14 (42.42%)	-	9 (40.91%)
Ethnicity				
White	-	30 (90.91%)	-	16 (72.73%)
Black	-	2 (6.06%)	-	5 (22.73%)
Latino	-	1 (3.03%)	-	0 (0.00%)
Other	-	0 (0.00%)	-	1 (4.55%)
Income <60,000/year	-	14 (42.42%)	-	6 (27.26%)
CBCL Internalizing (T)	69.03 (8.23)	-	65.00 (7.44)	-
CBCL Externalizing (T)	53.76 (10.05)	-	53.00 (11.25)	-
Total number of dx	3.33 (1.55)	-	2.68 (1.46)	-
Total anxiety dx	2.69 (1.26)	-	2.23 (1.02)	-
Total externalizing dx	.48 (.57)	-	.45 (.74)	-
Weeks of treatment	19.39 (3.55)	-	19.77 (4.75)	-

Order Scores and Adherence

Next, I turn to the order scores generated from the four different scoring conventions described above. In Table 3, below, the minimum, maximum, and mean values for each of the scoring conventions are presented. As was expected, the most liberal Content Only (CO) convention yielded the highest mean score, followed by the Skills Before Exposure (SBE), Practice Items in Prescribed Order (PIPO), and Practice Items in Prescribed Order at Prescribed Time (PIPOPT). Although individual case scores varied widely, mean scores were similar across each of the Scoring Conventions.

Table 3.

Minimum, Maximum and Mean Order Scores, by Convention (n = 33)

Scoring Convention	Minimum	Maximum	Mean (SD)
Content Only (CO)	3.00	4.60	3.74(.45)
Skills Before Exposure (SBE)	3.00	4.60	3.71(.46)
Practice Items in Prescribed Order (PIPO)	2.57	4.60	3.54(.54)
Practice Items in Prescribed Order at Prescribed Time (PIPOPT)	2.26	4.59	3.51(.55)

There are also similarities between scoring conventions reflected below in Table 4, which contains the Pearson bivariate correlations for the Scoring Conventions. All correlations are significant at $p < .00$ such that $r > .89$, indicating redundancy among the different systems.

Table 4.

Correlation of Order Score Conventions (n = 33)

	CO	SBE	PIPO	PIPOPT
Content Only (CO)	-			
Skills Before Exposure (SBE)	.98***	-		
Prescribed Items in Prescribed Order (PIPO)	.92***	.93***	-	
Prescribed Items in Prescribed Order at Prescribed Time (PIPOPT)	.90***	.91***	.89***	-

Also of interest was the frequency of "deviations" from order. As described above, the "skills-building" phase of treatment (i.e., building the FEAR plan) prescribes the order in which skills are taught. Deviation from this plan was observed in eleven cases (33.3%), as measured by the first session in which a skill was coded. There were four cases (12.1%) that one or more expected skill (e.g., cognitive, problem-solving, self-reward) was not present in the sessions

coded. Finally, there were five cases (15.2%) in which the exposure phase of treatment began early (i.e., before session 10).

Order Scores and Client Characteristics

I was also interested in assessing the relation between therapist adherence to order and the child's demographic and clinical characteristics. To do so, order scores were correlated with client characteristics, presented below in Table 5. Given the structure of the data, there were a number of considerations. First, because levels of measurement varied by child characteristic, we conducted both Pearson product moment correlations and Pearson point-biserial correlations, depending on the variables at hand. Second, because some child-level data are missing, individual Ns are presented for each analysis. Finally, because the reported ethnicity of the sample was overwhelmingly White (90.91%), ethnicity was not included among the correlations.

Table 5.

Correlation of Child and Treatment Characteristics with Order Scores by Convention

Characteristic	N	CO	SBE	PIPO	PIPOPT
Child Age	33	.04	.05	-.09	-.02
Gender[†]	33	-.14	-.15	-.06	-.18
Family Income[±]	29	-.05	-.02	.06	-.00
CBCL Internalizing (T-score)	33	-.01	.00	.11	.08
CBCL Externalizing (T-score)	33	-.10	-.11	-.07	-.04
Total number of diagnoses	33	-.29	-.23	-.19	-.22
Total anxiety diagnoses	33	-.32	-.27	-.22	-.28
Total externalizing diagnoses	33	-.05	.00	-.04	.01
Weeks of treatment	33	-.18	-.19	-.26	-.20

[†]Female coded positively.

[±]Dichotomous, with greater than $60,000/year coded positively.

Although no correlations emerged as statistically significant, we can see some notable effect size differences. Although the effect of age, gender, family income, and CBCL pre-treatment scores were largely negligible, diagnostic data and duration of treatment emerged as having a small effect ($r > .20$) across some scoring conventions, suggesting that as the duration of treatment lengthened and the number of diagnoses generally and anxiety diagnoses specifically increased, adherence to content and/or order decreased.

Item-Level Analysis

These adherence data can also help us understand what core CBT practices are being delivered more extensively. To provide a summary of such characteristics, I aggregated the adherence data at the CBAY-A item level. However, given the nature of adherence across the course of the case, it is expected that many items were coded as not present (extensiveness score = 1) in most sessions. To account for this distribution, we selected the maximum extensiveness score of each core model item of the CBAY-A from this sample of 33 cases and present the minimum, maximum, mode, and mean below in Table 6. For example, the minimum Psychoeducation-Anxiety value was 1.5, indicating that, in at least one case, the highest extensiveness score coded across the 12 included sessions was 1.5. In the case of Problem Solving and Self-Reward, the minimum value of 1.0 indicates that, in at least one case, these practices were not identified across the 12 included sessions.

Table 6.

Maximum Values of CBAY-A Items Within Each Case (n = 33)

Item	Minimum	Maximum	Mode	Mean (SD)
Psychoeducation-Anxiety	1.50	5.50	2.50	3.09 (.85)
Emotion Education	3.50	7.00	7.00	6.09 (.92)
Fear Ladder	3.00	6.50	4.50	4.45 (.88)
Relaxation	1.50	7.00	5.83	5.24 (1.29)
Cognitive-Anxiety	3.75	7.00	5.00	5.71 (.89)
Problem Solving	1.00	7.00	4.50	4.14 (1.87)
Self-Reward	1.00	6.00	5.00	4.52 (1.41)
Coping Plan	3.00	6.50	6.00	5.23 (1.04)
Exposure Composite	3.50	6.33	4.33	4.67 (.78)

From these data, we can see that some items were delivered far less extensively than others (e.g., Psychoeducation-Anxiety vs. Emotion Education) and that some items were somewhat limited in range, as no therapists in the sample delivered Psychoeducation-Anxiety, Fear Ladder, Self-Reward, Coping Plan, or the Exposure Composite "extensively" (extensiveness score = 7.0) over the course of their case(s). Also of note, there were some cases where only a low dose of individual CBT components were delivered (i.e., maximum values hovering around 3.0 and lower).

As order scores were generated using maximum item values delivered at a specific time, depending on the scoring convention, it is also helpful to see how maximum item-level scores correlate with the order scores themselves. Pearson product-moment coefficients are presented below in Table 7. Many items (Emotion Education, Cognitive-Anxiety, Problem-Solving, Self-Reward, Coping Plan, and Exposure Composite) significantly correlated with the order scores.

Table 7.

Correlation of Maximum CBAY-A Item Values with Order Scores (n = 33)

	CO	SBE	PIPO	PIPOPT
Psychoeducation	0.28	0.28	0.27	0.23
Emotion Education	0.43*	.45**	.53**	.60***
Fear Ladder	0.38*	.38*	0.33	0.25
Relaxation	0.14	0.10	0.22	0.16
Cognitive-Anxiety	0.42*	.53**	.52**	.50**
Problem Solving	0.62***	.58***	.49**	.56**
Self-Reward	0.45**	.51**	.38*	.36*
Coping Plan	0.62***	.60***	.48**	.48**
Exposure Composite	0.46**	.43*	.45**	.49**

Discussion

As described above, it is important for the field to expand its conceptualization of adherence to include the episode-level. This preliminary study has contributed to this effort by proposing four novel ways to measure adherence across the course of a case, using therapy process data from an efficacy trial of CBT for child anxiety as an example. Here, I turn to a discussion of the descriptive findings and the implications for the field at large there within.

Preliminary Findings

The four scoring systems were developed, conceptually, to represent an array of options that range from very liberal ("Content Only) to very conservative ("Practice Items in the Prescribed Order at the Prescribed Time"). Within this sample, the score distributions of each convention generally followed suit (e.g., mean scores decreased among more conservative scoring systems; see Table 3). The differences, however, were small, with all four mean scores falling within 0.25 points of one another. In Table 4, we can also see how highly correlated each

of the scoring systems are with one another, ranging from $r(31) = .89$ to $r(31) = .98$), further indicating duplicative measurement. This limited range may be the product of the sample itself: data were taken from an efficacy trial in which adherence to the Coping Cat protocol was closely monitored (Kendall et al., 2008). Further, the therapist sample was comprised primarily of graduate students, all of whom had extensive training in the intervention with the treatment developer. Had these scoring systems been applied to a larger sample of more diverse therapists (e.g., therapists working in the community with varying experience and supervision), a wider range of scores overall and meaningful differences between the different conventions may be captured.

Although the uniformity of scores is congruent with the efficacy trial sample, some may be surprised by relatively low adherence overall. Mean values hovered around 3.60 for all the scoring conventions; on the 1-7 extensiveness scale, this is indicative of adherence that just exceeds the "*somewhat* adherent" anchor at 3.0. Similarly, a number of cases that featured 'deviations' from the prescribed order were identified, such that the skills sessions were delivered out of order, skills were skipped, or exposure sessions began early. Given the importance of applying the independent variable consistently in controlled trials, these data suggest that the intervention being delivered is, as a whole, only moderately adherent to the manual. Although this may prove to be methodologically problematic, flexibility or deviation from adherence may not be detrimental to the 'quality' or effectiveness of treatment (e.g., Beidas & Kendall, 2010; Hogue et al., 2008). Further, these data are consistent with Park et al.'s (2015) findings of widespread deviation from prescribed order among therapists delivering both modular and linearly organized standard treatment in an effectiveness trial.

As I discuss above, there are many factors (e.g., child, therapist, and setting characteristics) that can influence the delivery of psychological treatments, making therapy process research particularly challenging. Here, the data presented in Table 5 suggest that diagnostic complexity (i.e., anxiety diagnoses, total diagnoses) was related to lower levels of adherence overall (i.e., Content Only scoring convention) and adherence to order specifically (i.e., other scoring conventions). Further, weeks of treatment negatively correlated to a modest degree with the more stringent scoring conditions, suggesting that as the duration of treatment lengthened, adherence to order fell.

It is unclear if prolonged treatment duration in this sample is indicative of limited family engagement (i.e., high no show rate, difficulty scheduling), or some other characteristic inherent to the child or family. Family engagement and barriers to treatment—both perceptual and concrete—have been shown to interfere with attendance (Mendez, Carpenter, LaForrett, & Cohen, 2009; Staudt, 2006), and, as Staudt (2006) describes in her model of treatment engagement, are integral to the delivery of treatment. Although the correlations in this sample were modest, treatment characteristics did relate to order adherence in the direction one would expect, such that increased clinical complexity and greater number of attendance disruptions were related to delivery of treatment in a less precise order.

Limitations

Perhaps the greatest limitation to the findings and method described above is the small sample size of n = 33. Despite the inclusion of data from hundreds of individual therapy sessions—all of which were double-coded—measurement at the episode-level is necessarily reductive and yields are subsequently low. Further, the processing that was necessary to ensure

data yielded meaningful order scores (e.g., imputing missing data, calculating order scores themselves) was substantial.

Missing data also posed a significant problem. Because complete case data were necessary to calculate order in the way proposed, data imputation was unavoidable. Although I attempted to minimize the influence of these missing values by selecting cases for which we had most data and computed imputation values based upon whole cases in the dataset, there is no way to know if and how this may have affected the calculation of order scores. Similarly, by excluding the first, last, and parent-only sessions from our analyses, the ability to provide a complete accounting of what was delivered across the case is limited.

Therapy process models (e.g., McLeod et al., 2011) often include a number of characteristics present before and during treatment that may affect the delivery and/or receipt of therapeutic interventions. Although there was considerable child-specific information (e.g., age, gender, ethnicity, income) available, therapist-specific factors (e.g., theoretical orientation, attitudes toward evidence-based practice) and relational factors (e.g., therapeutic alliance, client involvement) were unavailable and excluded from analyses. The inclusion of such information may provide a more nuanced description of how and why therapists deviated from prescribed order.

Finally, given the lack of discussion and consensus about this topic in the field, we were forced to develop our order scoring conventions and method with a narrow scope (i.e., the Coping Cat intervention) and without much guidance. We attempted to base our scoring conventions on theory (e.g., Skills Before Exposure scoring convention) or empirical information (e.g., coding the Coping Cat manual as a comparator for the scoring conventions). In using a Coping Cat trial as an example of how we might measure order, the specific scoring rules are

only applicable to the Coping Cat intervention; should others attempt to apply such measurement to other brands of therapy, they will first need to identify or develop their own comparator. Park et al. (2015) describe similar comparator procedures in their comparison of practice arrangement in modular and linear treatments (i.e., using the algorithm or manual to create a template of expected practice content), but scored individual sessions based upon what was delivered in previous sessions, rather than the session number. As the field continues to explore the structure of psychosocial treatments, it will be important to determine a method for measuring treatment integrity at the episode-level.

Future Directions and Applications

Although I have presented a fairly nuanced description of treatment delivery within one arm of an RCT, the true product of this study is the novel method by which these results were ascertained. The conceptualization of treatment integrity generally—and adherence specifically—as occurring across a case has raised a number of questions for those in the fields of efficacy and/or effectiveness research, therapy process research, and dissemination and implementation science. Here, we discuss possible applications of such measurement within each of these domains.

Efficacy and/or effectiveness research. Perhaps the most obvious application of measuring adherence across a case is providing another check of the independent variable as it is delivered. As has been demonstrated here, researchers may be able to use treatment manuals to develop treatment-specific comparators to compare therapy process data. If these data are collected and processed quickly, therapists and their supervisors may be able to use them to guide future sessions, further reinforcing the delivery of a consistent independent variable across cases.

As modular treatments—in which therapists make decisions about what 'modules' (e.g., “Relaxation”, “Psychoeducation”) they deliver based upon specific rules or flowcharts—continue to be developed and tested (e.g., Chorpita & Weisz, 2009), the field seems poised to explore more complex adherence to order measurement. The definition of adherence to these treatments may differ dramatically from case to case, depending largely upon clinical judgment and assessment data (e.g., symptom ratings) that are available to the clinician. Park and colleagues (2015) leveraged the clinician-completed measure of adherence collected every session (i.e., the Consultation Record) and developed comparators based on the arrangement of modular and standard protocols prescribed in decisional flowcharts and treatment manuals, respectively.

As the field continues to explore the benefits of adaptable treatment protocols that promote therapist flexibility, it will be increasingly important to use valid and reliable measures of treatment integrity as a way to maintain the careful balance between structure and flexibility while maximizing effectiveness.

Therapy process research. As mentioned briefly above, the inclusion of additional therapy process data (e.g., therapist competence, treatment differentiation, relational measures, therapist characteristics, session-by-session treatment progress) may help inform our understanding of *why* therapists deviate from a given manual. For example, some have shown a temporal relation between level of therapist-client alliance and subsequent therapist adherence (Chu & Kendall, 2009). As we have demonstrated, however, measurement of adherence most often at the event-level, neglecting to account for what has been delivered in previous sessions or how treatment components fit within adherence across the case. Incorporating measurement of episode-level adherence into such therapy process models, in conjunction with the many pre-

treatment and process variables that influence and inform treatment, could help the field develop a more complete understanding of why and how treatment occurs over time.

Dissemination and implementation science. Although the present study has focused on the delivery of services in a university training clinic, children and families access mental health care in many different settings (e.g., community clinics, juvenile justice, schools, primary care, child welfare; Farmer, Burns, Phillips, Angold, & Costello, 2003). Families may have differential access to these services due, in part, to various child-, family-, and community-level factors (e.g., family insurance status, proximity to services, community funding for services; e.g., Kataoka, Zhang, & Wells, 2002), resulting in unique populations served by each service setting. Further, each of these service systems may have unique challenges, strengths, and goals that affect how mental health services are delivered (e.g., availability of training opportunities, administrative support, organizational structures; see Fixsen, Naoom, Blase, Friedman, & Wallace, 2005, for a discussion). As the field attempts to improve practice within such systems, measures of treatment integrity can help us understand how the structure of therapy may differ across setting, population, target problem, etc. To this end, measuring treatment integrity, generally, and adherence, specifically, across the course of a case may provide a way to document any setting-specific differences that can then (1) inform how systems implement treatments, or (2) help researchers develop novel treatments for specific settings (e.g., managing and/or anticipating trends of therapist behavior).

Ultimately, it would seem that conceptualizing and measuring treatment integrity at the episode-level may provide a more complete picture of how therapy is delivered compared to measurement at the individual session level. Here, I have developed one way of measuring adherence across the course of a case that yielded a rich and nuanced description of I-CBT as

delivered in the Coping Cat trial. Although more research is certainly needed to refine and apply such measurement to other settings, the accurate quantification and documentation of how therapists deliver treatment, from start to finish, has the potential to enhance the field in many ways.

List of References

Aarons, G. A. (2004). Mental health provider attitudes toward adoption of evidence-based practice: The Evidence-Based Practice Attitude Scale (EBPAS). *Mental Health Services Research, 6*(2), 61-74.

Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles.* Burlington, VT: University of Vermont, Department of Psychiatry.

Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes towards psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology, 68,* 331-339.

Albano, A. M., & Kendall, P. C. (2002). Cognitive behavioural therapy for children and adolescents with anxiety disorders: Clinical research advances. *International Review of Psychiatry, 14,* 129-134. doi:10.1080/09540260220132644

Allison, P. D. (2002). *Missing data.* Sage University Papers Series on Quantitative Applications in the Social Sciences, 07*136. Thousand Oaks, CA: Sage.

American Psychiatric Association. (1980). *Diagnostic & statistical manual of mental disorders* (3rd ed). Washington, DC: Author.

Anderson, J. C., Williams, S., McGee, R., & Silva, P. A. (1987). DSM-III disorders in preadolescent children: Prevalence in a large sample from the general population. *Archives of General Psychiatry, 44,* 69-76.

Barber, J. P., Gallop, R., Crits-Cristolph, P., Frank, A., Thase, M. E., Weiss, R. D., & Gibbons, M. B. C. (2006). The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counseling: Results from the National Institute Drug Abuse Collaborative Cocaine Treatment Study. *Psychotherapy Research, 16,* 229-240. doi: 10.1080/10503300500288951

Barrington, J., Prior, M., Richardson, M., & Allen, K. (2005). Effectiveness of CBT versus standard treatment for childhood anxiety disorders in a community clinic setting. *Behaviour Change, 22,* 29-43. doi: http://dx.doi.org/10.1375/bech.22.1.29.66786

Baumann, B. L., Kolko, D. J., Collins, K., & Herschell, A. D. (2006). Understanding practitioners' characteristics and perspectives prior to the dissemination of evidence-based intervention. *Child Abuse & Neglect, 30,* 771-787.

Beidas, R. S., Benjamin, C. L., Puleo, C. M., Edmunds, J. M., & Kendall, P. C. (2010). Flexible applications of the Coping Cat program for anxious youth. *Cognitive and Behavioral Practice, 17,* 142-153.

Beidas, R. S., Edmunds, J., Ditty, M., Watkins, J., Walsh, L., Marcus, S., & Kendall, P. (2013). Are inner context factors related to implementation outcomes in cognitive-behavioral therapy for youth anxiety? *Administration and Policy in Mental Health and Mental Health Services.* Advance online publication. doi: 10.1007/s10488-013-0529-x

Bein, E., Anderson, T., Strupp, H. H., Henry, W. P., Schact, T. E., Binder, J. L., & Butler, S. F. (2000). The effects of training in Time-Limited Dynamic Psychotherapy: Changes in therapeutic outcome. *Psychotherapy Research, 10,* 119-132. doi: 10.1093/ptr/10.2.119

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., … Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology, 23,* 443-451. doi:10.1037/0278-6133.23.5.443

Bernstein, G. A., & Borchardt, C. M. (1991). Anxiety disorders of childhood and adolescence: A critical review. *Journal of the American Academy of Child & Adolescent Psychiatry, 30,* 519-532.

Boswell, J. F., Gallagher, M. W., Sauer-Zavala, S. E., Bullis, J., Gorman, J. M., Shear, M. K., … Barlow, D. H. (2013). Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for panic disorder. *Journal of Consulting and Clinical Psychology, 81,* 443-454.

Brookman-Frazee, L., Garland, A. F., Taylor, R., & Zoffness, R. (2009). Therapists' attitudes towards psychotherapeutic strategies in community-based psychotherapy with children with disruptive behavior problems. *Administration and Policy in Mental Health and Mental Health Services, 36,* 1-12.

Brown, R. C., Southam-Gerow, M. A., McLeod, B. D., Bair, C. E., & Wheat, E. (2012). *Scoring manual for the Common Factor Therapist Competence Scale.* Unpublished manual. Psychology Department, Virginia Commonwealth University, Richmond, VA.

Burns, B. J., Hoagwood, K., & Mrazek, P. J. (1999). Effective treatment for mental disorders in children and adolescents. *Clinical Child and Family Psychology Review, 2,* 199-254).

Child and Adolescent Mental Health Division. (2003). *Instructions and codebook for provider monthly summaries.* Honolulu, HI: Hawaii Department of Child and Adolescent Mental Health Division.

Chorpita, B. F., Daleiden, E. L., Ebesutani, C., Young, J., Becker, K. D., Nakamura, B. J., … Starace, N. (2011). Evidence-based treatments for children and adolescents: An updated

review of indicators of efficacy and effectiveness. *Clinical Psychology: Science and Practice, 18*, 153-171.

Chorpita, B. F., & Weisz, J. R. MATCH-ADTC: Modular approach to therapy for children with anxiety, depression, trauma, or conduct problems. *PracticeWise, LLC*: Satellite Beach, FL.

Chu, B. C., & Kendall, P. C. (2009). Therapist responsiveness to child engagement: Flexibility within manual-based CBT for anxious youth. *Journal of Clinical Psychology, 65,* 736-754. doi:10.1002/jclp.20582

Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284-290.

Connor-Smith, J. K., & Compas, B. E. (2003). Analogue measures of DSM-IV mood and anxiety disorders based on behavior checklists. *Journal of Psychopathology and Behavioral Assessment, 25*, 37-48. doi: 10.1023/A:1022299903358

Daleiden, E. L., & Chorpita, B. F. (2005). From data to wisdom: Quality improvement strategies supporting large-scale implantation of evidence-based services. *Child and Adolescent Psychiatric Clinics of North America, 14,* 329-349. doi:10.1016/j.chc.2004.11.002

Dobson, K. S., & Singer, A. R. (2005). Definitional and practical issues in the assessment of treatment integrity. *Clinical Psychology: Science and Practice, 12,* 384-387. doi:10.1093/clipsy/bpi046

Epstein, A. M., Lee, T. H., & Hamel, M. B. (2004). Paying physicians for high-quality care. *New England Journal of Medicine, 350*, 406-410.

Erhardt, K. E., Barnett, D. W., Lentz, F. E., Jr., Stollar, S. A., & Reifan, L. H. (1996). Innovative methodology in ecological consultation: Use scripts to promote treatment acceptability and integrity. *School Psychology Quarterly, 11*, 149-168.

Eyberg, S. M., Boggs, S. R., & Algina, J. (1995). Parent-child interaction therapy: A psychosocial model for the treatment of young children with conduct problem behavior and their families. *Psychopharmacology Bulletin, 31*, 83-91.

Farmer, E. M. Z, Burns, B. J., Phillips, S. D., Angold, A., & Costello, J. E. (2003). Pathways into and through mental health services for children and adolescents. *Psychiatric Services, 54*, 60-66.

Fergusson, D. M., Horwood, L. J., & Lynskey, M. T. (1993). Prevalence and comorbidity of DSM-III-R diagnoses in a birth cohort of 15 year olds. *Journal of the American Academy of Child & Adolescent Psychiatry, 32*, 1127-1134. doi: 10.1097/00004583-199311000-00004

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M. & Wallace, F. (2005). *Implementation research: A synthesis of the literature.* Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Flannery-Schroeder, E. C., & Kendall, P. C. (2000). Group and individual cognitive-behavioral treatments for youth with anxiety disorders: A randomized clinical trial. *Cognitive Therapy and Research, 24*, 251-278. doi: 10.1023/A:1005500219286

Forehand, R., Dorsey, S., Jones, D. J., Long, N., & McMahon, R. J. (2010). Adherence and flexibility: They can (and do) coexist! *Clinical Psychology: Science and Practice, 17,* 258-264.

Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65*, 821-833. doi: 10.1037/0022-006X.65.5.821

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (2009). *Multisystemic therapy for antisocial behavior in children and adolescents* (2nd ed.). New York, NY: Guilford Press.

Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A., & Liddle, H. A. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology, 76*, 544-555. doi: 10.1037/0022-006X.76.4.544

Huey, S. J., Henggeler, S. W., Brondino, M. J., & Pickrel, S. G. (2000). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology, 68*, 451-467. doi: 10.1037/0022-006X.68.3.451

Jones, K. M., Wickstrom, K. F., & Friman, P. C. (1997). The effects of observational feedback on treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 12*, 316-326.

Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet need for mental health care among US children: Variation by ethnicity and insurance status. *American Journal of Psychiatry, 159*, 1548–1555.

Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston, MA: Allyn & Bacon.

Kendall, P. C. (1993). Cognitive-behavioral therapies with youth: Guiding theory, current status, and emerging developments. *Journal of Consulting and Clinical Psychology, 61,* 235-247.

Kendall, P. C. (1994). Treating anxiety disorders in children: Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology, 62*, 100-110.

Kendall, P. C., Flannery-Schroeder, E., Panichelli-Mindel, S. M., Southam-Gerow, M., Henin, A., & Warman, M. (1997). Therapy for youths with anxiety disorders: A second randomized clinical trial. *Journal of Consulting and Clinical Psychology, 65*, 366-380.

Kendall, P. C., & Beidas, R. S. (2007). Smoothing the trial for dissemination of evidence-based practices for youth: Flexibility within fidelity. *Professional Psychology: Research and Practice, 38,* 13-20. doi:10.1037/0735-7028.38.1.13

Kendall, P. C., Furr, J. M., & Podell, J. L. (2010). Child-focused treatment of anxiety. In J. R. Weisz and A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (2nd ed.) (pp. 45-60). New York, NY: Guilford Press.

Kendall, P. C., Gosch, E., Furr, J. M., & Sood, E. (2008). Flexibility within fidelity. *Journal of the American Academy of Child & Adolescent Psychiatry, 47,* 987-993. doi: 10.1097/CHI.0b013e31817eed2f

Kendall, P. C., & Hedtke, K. (2006). *Cognitive-behavioral therapy for anxious children: Therapist manual* (3rd ed). Ardmore, PA: Workbook Publishing.

Kendall, P. C., Hudson, J. L., Gosch, E., Flannery-Schroeder, E., & Suveg, C. (2008). Cognitive-behavioral therapy for anxiety disordered youth: A randomized clinical trial evaluating child and family modalities. *Journal of Consulting and Clinical Psychology*, *76*, 282-297. doi: 10.1037/0022-006X.76.2.282

Kendall, P. C., Kane, M., Howard, B., & Siqueland, L. (1990). C*ognitive behavioral therapy for anxious children: Treatment manual*. (Available from Phillip C. Kendall, Department of Psychology, Temple University, Philadelphia, PA, 19122).

Kendall, P. C., & Southam-Gerow, M. A. (1996). Long-term follow-up of a cognitive-behavioral therapy for anxiety-disordered youth. *Journal of Consulting and Clinical Psychology, 64*, 724-730. doi: 10.1037/0022-006X.64.4.724

Lundgren, L., Amodeo, M., Chassler, D., Krull, I., & Sullivan, L. (2013). Organizational readiness for change in community-based addiction treatment programs and adherence in implementing evidence-based practices: A national study. *Journal of Substance Abuse Treatment, 45*, 457-465.

Lutzker, J. R., & Bigelow, K. M. (2002*). Reducing child maltreatment: A guidebook for parent services.* New York, NY: Guilford Press.

Margolin, G., Oliver, P. H., Gordis, E. B., O'Hearn, H. G., Medina, A. M., Ghosh, C. M., & Morland, L. (1998). The nuts and bolts of behavioral observation of marital and family

interaction. *Clinical Child and Family Review, 1*, 195-213. doi: 10.1023/A:1022608117322

Mash, E. J., & Barkley, R. A. (Eds.). (2003). *Child psychopathology* (2nd ed.). New York, NY: Guilford Press.

McClure, E. B., & Pine, D. S. (2006). Social anxiety and emotion regulation A model for developmental psychopathology perspectives on anxiety disorders. In D. Cicchetti and D. J. Cohen (Eds.), *Developmental psychopathology, Vol. 3: Risk disorder, and adaptation* (2nd ed.) (pp. 470-502). Hoboken, NJ: John Wiley & Sons, Inc.

McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodríguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health care. *Clinical Psychology: Science and Practice, 44*, 525-536. doi: 10.1111/cpsp.12020

McLeod, B. D., Southam-Gerow, M. A., & Weisz, J. R. (2009). Conceptual and methodological issues in treatment integrity measurement. School Psychology Review, 38, 541-546.

McLeod, B. D., & Weisz, J. R. (2010). The Therapy Process Observational Coding System for Child Psychotherapy Strategies scale. *Journal of Clinical Child and Adolescent Psychology, 39*, 436-443. doi: 10.1080/15374411003691750

Mendez, J. L., Carpenter, J. L., LaForett, D. R., & Cohen, J. S. (2009). Parental engagement and barriers to participation in a community-based preventive intervention. *American Journal of Community Psychology, 44*, 1-14.

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*, 247-266. doi: 10.1016/0272-7358(91)90103-2

Park, A. L., Chorpita, B. F., Regan, J., Weisz, J. R., & The Research Network on Youth Mental Health. (2015). Integrity of evidence-based practice : Are providers modifying practice content or practice sequencing? *Administration and Policy in Mental Health, 42*, 186-196.

Perepletchikova, F. (2009). Treatment integrity and differential treatment effects. *Clinical Psychology: Science and Practice, 16,* 379-382.

Perepletchikova, F. (2011). On the topic of treatment integrity. *Clinical Psychology: Science and Practice, 18,* 148-153.

Perepletchikova, F., Hilt, L. M., Chereji, E., and Kazdin, A. E. (2009). Barriers to implementing treatment integrity procedures: Survey of treatment outcome researchers. *Journal of Consulting and Clinical Psychology, 77,* 212-218.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice, 12,* 365-383. doi:10.1093/clinpsy/bpi045

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology, 75,* 829-841. doi: 10.1037/0022-006X.75.6.829

Peterson, L., Homer, A. L., and Wanderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15*, 477-492. doi: 10.1901/jaba.1982.15-477

Regan, J., Daleiden, E. L., & Chorpita, B. F. (2013). Integrity in mental health systems: An expanded framework for managing uncertainty in clinical care. *Clinical Psychology: Science and Practice, 20,* 78-98. doi: 10.1111/cpsp.12024

Rosenthal, M. B., Fernandopulle, R., Song, H. R., & Landon, B. (2004). Paying for quality: Providers' incentives for quality improvement. *Health Affairs, 23*, 127-141.

Sanders, M. R. (1999). Triple P – Positive Parenting Program: Towards an empirically validated multilevel parenting and family support strategy for the prevention of behavior and emotional problems in children. *Clinical Child and Family Psychology Review, 2,* 71-90. doi: 10.1023/A:1021843613840

Schoenwald, S. K. (2008). Toward evidence-based transport of evidence-based treatments: MST as an example. *Journal of Child and Adolescent Substance Abuse, 17*, 69-91.

Schoenwald, S. K. (2011). It's a bird, it's a plane, it's … fidelity measurement in the real world. *Clinical Psychology: Science and Practice, 18,* 142-147.

Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research, 38,* 32-43. doi:10.1007/s10488-010-0321-0

Schoenwald, S. K., Garland, A. F., Southam-Gerow, M. A., Chorpita, B. F., & Chapman, J. E. (2011). Adherence measurement in treatments for disruptive behavior disorders: Pursuing clear vision through varied lenses. *Clinical Psychology: Science and Practice, 18,* 331-341.

Schoenwald, S. K., Sheidow, A. J., & Letourneau, E. J. (2004). Toward effective quality assurance in evidence-based practice: Links between expert consultation, therapist fidelity, and child outcomes. *Journal of Clinical Child and Adolescent Psychology, 33,* 94-104.

Sheidow, A. J., Donohue, B. C., Hill, H. H., Henggeler, S. W., & Ford, J. D. (2008). Development of an audiotape review system for supporting adherence to an evidence-based treatment. *Professional Psychology: Research and Practice, 39,* 553-560. doi: 10.1037/0735-7028.39.5.553

Silverman, W. K., & Berman, S. L. (2001). Psychosocial interventions for anxiety disorders in children: Status and future directions. In W. K. Silverman and P. D. A. Treffers (Eds.) *Anxiety disorders in children and adolescents: Research, assessment, and intervention* (pp. 313-334). New York, NY: Cambridge University Press.

Southam-Gerow, M. A., McLeod, B. M., Arnold, C. C., Rodriguez, A., Cox, J. R., Reise, S. P. … Kendall, P. C. (in press). Initial development of a treatment adherence measure for cognitive-behavioral therapy for child anxiety. *Psychological Assessment.*

Southam-Gerow, M. A., Weisz, J. R., Chu, B. C., McLeod, B. D., Gordis, E. B., & Connor-Smith, J. K. (2010). Does cognitive behavioral therapy for youth anxiety outperform usual care in community clinics? An initial effectiveness test. *Journal of the American Academy of Child & Adolescent Psychiatry, 49,* 1043-1052.

Southam-Gerow, M. A., & McLeod, B. D. (2011). *Scoring manual for the CBT for Youth Anxiety Therapist Adherence Scale (ADH-CBT-YA).* Unpublished manual. Psychology Department, Virginia Commonwealth University, Richmond, VA.

Southam-Gerow, M. A., & McLeod, B. D. (2013). *Scoring manual for the CBT for Youth Anxiety (COMP-CBT-YA).* Unpublished manual. Psychology Department, Virginia Commonwealth University, Richmond, VA.

Staudt, M. (2007). Treatment engagement with caregivers of at-risk children: Gaps in research and conceptualization. *Journal of Child and Family Studies, 16,* 183-196.

Sterling-Turner, H. E., Watson, T. S., & Moore, J. W. (2002). The effects of direct training and treatment integrity on treatment outcomes in school consultation. *School Psychology Quarterly, 17,* 47-77. doi: 10.1521/scpq.17.1.47.19906

Suveg, C., Kendall, P. C., Comer, J. S., & Robin, J. (2006). Emotion-focused cognitive-behavioral therapy for anxious youth: A multiple-baseline evaluation. *Journal of Contemporary Psychotherapy,* 36, 77-85. doi: 10.1007/s10879-006-9010-4

U.S. Public Health Service. (2000). *Report of the surgeon general's conference on children's mental health: A national action agenda.* Washington, DC: Department of Health and Human Services.

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology,* 620-630. doi: 10.1037/0022-006X.61.4.620

Ward, A. M., Regan, J., Chorpita, B. F., Starace, N., Rodriguez, A. … The Research Network on Youth Mental Health. (2013). Tracking evidence based practice with youth : Validity of the MATCH and standard manual Consultation Records. *Journal of Clinical Child & Adolescent Psychology, 42,* 44-55.

Weersing, V. R., Weisz, J. R., & Donenberg, G. R. (2002). Development of the Therapy Procedures Checklist: A therapist-report measure of technique use in child and adolescent treatment. *Journal of Clinical Child and Adolescent Psychology, 31,* 168-180. doi: 10.1207/153744202753604458

Weisman, A., Tompson, M. C., Okazaki, S., Gregory, J., Goldstein, M. J., Rea, M., & Miklowitz, D. J. (2002). Clinicians' fidelity to a manual-based family treatment as a predictorof the one-year course of bipolar disorder. *Family Process, 41,* 43-52. doi: 10.1111/j.1545-5300.2002.40102000123.x

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., … Research Network on Youth Mental Health. (2012). Testing standard and modular designs for psychotherapy for youth depression, anxiety, and conduct problems: A randomized effectiveness trial. *Archives of General Psychiatry, 69,* 274-282.

Whitaker, D. J., Ryan, K. A., Wild, R. C., Self-Brown, S., Lutzker, J. R., Shanley, J. R. … Hodges, A. E. (2012). Initial implementation indicators from a statewide rollout of SafeCare within a child welfare system. *Child Maltreatment, 17,* 96-101. doi: 10.1177/1077559511430722

World Health Organization (WHO). (2005). *Mental health policy and service guidance package: Child and adolescent mental health policies and plans.* Geneva: Author.

Appendix A

Coping Cat Session Guide

Session	Corresponding CBAY-A Items
1. Building Rapport and Treatment Orientation	Excluded
2. Identifying Anxious Feelings	Emotion Education Fear Ladder
3. Identifying Somatic Responses to Anxiety	Emotion Education Psychoeducation
4. Parent Session	Excluded
5. Relaxation Training	Relaxation Emotion Education
6. Identifying Anxious Self-Talk and Learning to Challenge Thoughts	Cognitive-Anxiety Coping Plan
7. Reviewing Anxious and Coping Self-Talk and Developing Problem Solving Skills	Coping Plan Problem Solving
8. Introducing Self-Evaluation and Self-Reward and Reviewing Skills Already Learned	Coping Plan Self-Reward
9. Parent Session	Excluded
10. Practicing in Low Anxiety-Provoking Situations Using Exposure Tasks	Coping Plan Exposure Composite
11. Practicing in Low Anxiety-Provoking Situations Using Exposure Tasks	Coping Plan Exposure Composite
12. Practicing in Moderately Anxiety-Provoking Situations Using Exposure Tasks	Coping Plan Exposure Composite
13. Practicing in Moderately Anxiety-Provoking Situations Using Exposure Tasks	Coping Plan Exposure Composite
14. Practicing in High Anxiety-Provoking Situations Using Exposure Tasks	Coping Plan Exposure Composite
15. Practicing in High Anxiety-Provoking Situations Using Exposure Tasks	Coping Plan Exposure Composite
16. Practicing in High Anxiety-Provoking Situations Using Exposure Tasks, and Terminating Treatment	Excluded

Vita

Julia Revillion Cox was born on August 27, 1987, in Northridge, California, and is a citizen of the United States and Brazil. She graduated from Granada Hills Charter High School, Granada Hills, California in 2005. She received her Bachelor of Science in Psychology with departmental honors from Seattle University in 2009 and subsequently worked as a Research Coordinator at the University of Washington for three years.