Evaluating Case Conceptualizations in Psychotherapy Reports: Links to Therapy Outcome and the Alliance

by

Kevin Capobianco

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Psychology

Waterloo, Ontario, Canada, 2015

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Case conceptualizations are a key component of cognitive behavioural therapy (Beck, 1995; Persons, 2005; Needleman, 1999). Despite the theoretical importance of case conceptualizations, the question of whether they actually improve therapy outcomes is relatively unexplored (Bieling & Kuyken, 2003). Additionally, case conceptualizations may have other important effects on therapy, such as by increasing client engagement and improving the therapeutic alliance (Nattrass, Kellett, Hardy, & Ricketts, 2014). Utilizing two approaches (Collaborative Case Conceptualization Rating Scale; Padesky, Kuyken, & Dudley, 2011; Case Formulation Content Coding Method; Eells, Kendjelic, & Lucas, 1998) for evaluating case conceptualization quality, this project examines these possible inter-relations. The project involved a trained team of coders rating case conceptualizations found within psychotherapy reports generated at two stages of therapy. The first reports coded were written following the completion of the assessment phase of therapy, and the later reports were written upon client discharge from therapy. In Study 1 the comparative reliabilities in coding achieved across the two methods utilized is discussed, as are the strengths and weaknesses in coding and their possible causes. Study 2 reports the results obtained upon examining the relationship between case conceptualization quality, the therapeutic alliance, and treatment outcome. Results of Study 2 suggest that for more complicated or impaired client cases, therapists produce higher quality conceptualizations at assessment but these generally do not predict therapy outcome or the overall alliance between therapist and client. It was found that the quality of conceptualizations within discharge reports were associated with more positive therapy outcomes. Case conceptualization quality at discharge was also positively associated with the therapeutic alliance.

Acknowledgements

My sincere thanks go to my supervisor, Dr. Jonathan Oakman, whose encouragement and guidance were greatly appreciated over the course of this project. Similarly, Dr. Walter Mittelstaedt, who continues to be involved with the Psychological Intervention Research Team (PIRT) despite his new and demanding position, must be thanked. The work of the PIRT research assistants must also be recognized, without their dedication and enthusiasm this research could not have been done. The Social Sciences and Humanities Research Council of Canada has also been instrumental by providing the funding that supports our lab and our research. I also extend my appreciation to Dr. Christine Purdon and Dr. David Moscovitch who graciously have taken on the role of readers for this thesis and offered their insightful and constructive comments. Finally, I owe a huge thank you to my friends and family for their unwavering support while I worked towards the completion of this project.

Table of Contents

List of Tables

Case Conceptualization Research and Theory: A Literature Review

The process of psychotherapy involves the exchange of a great deal of information between client and therapist. This necessitates some method of synthesizing and distilling the information gathered in therapy sessions into a form that is more accessible to clinicians and can guide the direction of therapy. This process of integrating the experiences, background, symptoms, and goals of clients represents the development of a case conceptualization (also known as a case formulation). Generally a case conceptualization contains the working hypotheses for the causes of a client's difficulties, both long term and more acutely, and highlights the key features of a client's distress and impairment (Kuyken, Fothergill, Musa, & Chadwick, 2005).

The organizational function of case conceptualizations may enhance psychotherapy as it produces a concise and accessible theory for the symptoms and problems a client and therapist wish to tackle (Benjamin, 2003). Moreover, many other benefits of utilizing case conceptualizations in psychotherapy have been proposed, most of which span psychotherapeutic orientations. Authors within the case conceptualization literature propose that case conceptualizations promote insight and engagement in the client, help to focus and prioritize which interventions to deploy, help therapists anticipate possible problems in therapy, simplify complex problems, and validate and normalize a patient's presenting issues (Beck, 1995; Eells, 2011; Needleman, 1999; Persons & Tompkins, 1997).

Case conceptualizations in psychotherapy can also be seen as an alternative to what some clinicians perceive as a problematic trend towards more diagnosis-guided, rigid, and standardized approaches to psychotherapy. Some feel that diagnoses alone may not provide

enough background and substance to guide therapy in an ideal manner (Restifo, 2011). This issue may be most apparent when considering standardized or manualized therapies. These therapy approaches are typically empirically supported methods for treating specific disorders and often include session guidelines and targets for therapy. The growth of empirically supported therapies (ESTs) is generally seen as a major advance in clinical psychological work, as many agree that treatment efficacy should be demonstrated within structured research designs (Chambless & Hollon, 1998). Despite the obvious benefits of research backed psychotherapies, some clinicians believe that information valuable to guiding treatment, such as which symptoms or diagnoses are most impairing the patient, or which may be exacerbating others, or what changes and techniques would most benefit the client, lie outside of the realm of a diagnosis alone, which is often the primary metric for selecting and validating manualized interventions (Persons, 1991; Persons, 2006).

Persons' (1991) commentary also highlights a potentially limiting effect of an emphasis on empirically supported therapies. The crux of this issue is that manualized therapies often encourage some clinical flexibility in their use; however, this results in poor generalizability of the research findings from the development of these ESTs to clinical practice as it regularly occurs. If a particular therapy approach is demonstrated to be effective in a randomized control trial, where the population of clients and the treatment adherence of the therapists are both tightly controlled, it may not generalize to therapy as actually practiced. Thus, while research continues to find support for treatments of specific disorders (Aston, 2009; Chambless et al., 1998), it may fail to shed light on therapy outcomes in naturalistic settings. Those who seek mental health treatments from psychologists and other providers often have comorbid diagnoses, or do not fit into a diagnostic category cleanly (Adam, Meinlschmidt, Gloster, & Lieb, 2012; Newman,

Przeworski, Fisher, & Borokovec, 2010), threatening the generalizability of results obtained from studies of standardized treatments (Persons, 2005).

For this reason case conceptualizations have been seen as one way to incorporate an empirical and theory driven approach to therapy into areas of clinical practice where research has not yet been conducted. Similarly, case conceptualizations may be most useful in situations where the complexity and uniqueness of a client's case limits a therapist's ability to draw upon EST research to find appropriate treatment options or where ESTs offer some flexibility in their deployment and clinicians must make decisions on which interventions to deploy and at which point in therapy to deploy them (Mumma & Smith, 2001). It may be that case conceptualizations offer a method for ensuring an empirically derived intervention while also providing the flexibility needed for the treatment of ideographic presentations that do not clearly match with the clients found in RCTs (Sim, Gwee, & Bateman, 2005).

Given that case conceptualizations appear to be widely held as valuable for enhancing therapy it is unsurprising that they can be found within psychodynamic (Crits-Christoph, Cooper, & Luborsky, 1988), behavioural (Haynes & Williams, 2003), and cognitive behavioral therapy traditions (Beck, 1995), among others. Within the realm of cognitive behavioural therapy (CBT), case conceptualizations have been described as a key competency or first principle (Beck, 1995; Persons 1989; Needleman, 1999) that lies at the heart of practice. As outlined by Kuyken and colleagues (2005) the purpose of cognitive behavioural case conceptualizations can be described as follows:

> *For any particular case of CBT practice, formulation is the bridge between practice, theory, and research. It is the crucible where the individual*

> *particularities of a given case, relevant theory and research synthesize into an understanding of the person's presenting issues in CBT terms that informs the intervention.* (pp. 1188)

Although many authors propose benefits for utilizing case conceptualizations, few studies have examined the actual impacts of case conceptualization on treatment outcomes (Bieling & Kuyken, 2003). This is an issue for those authors who suggest that case conceptualization guided psychotherapy can fill the empirical gap between the ideographic treatments offered in naturalistic settings and the standardized treatments offered within RCTs. We will return to the little research that has been done in regards to case conceptualizations and treatment outcomes; however, research addressing more basic questions regarding conceptualizations will be addressed first.

Reliability in a psychological construct or test is often seen as the essential bedrock upon which later examinations of validity must lie (Shrout & Lane, 2012). Given the foundational importance of establishing the reliability of any given construct, it is unsurprising that a sizable body of research has explored the reliability of case conceptualizations. In a seminal study by Crits-Christoph, Cooper, & Luborsky (1988) the Core Conflictual Relationship Themes (CCRT) method produced reliable conceptualizations across judges. Barber and Crits-Christoph (1993) later reviewed reliability research across several other methods for generating and evaluating psychodynamic formulations. In this review, they conclude that although many of the results reported are preliminary, they are encouraging, in that the main themes for maladaptive interpersonal patterns appear as though they can be reliably judged across raters. More recent research into CBT conceptualizations promisingly suggests that a degree of reliability can also be achieved across CBT oriented clinicians.

In one study examining the reliability of CBT conceptualizations, the case conceptualization diagram (CCD) method (Beck, 1995) was utilized and assessed by Kuyken, Fothergill, Musa, and Chadwick (2005). Within their study CCDs were generated following a case presentation as part of a training workshop on the CCD method. The clinician participants varied in their professional backgrounds, with clinical psychologists, psychiatric nurses, counsellors, and pre-qualification students comprising the largest groups. The independently generated CCDs were compared against each other and to a benchmark CCD provided by Judith Beck. Results indicated that for formulation categories of relevant childhood data, core beliefs, and compensatory strategies, the raters showed high levels of agreement in including specific elements within their formulations. On the level of dysfunctional assumptions, however, agreement across raters was lower. Conclusions drawn from this research were that given appropriate training and a structured method for developing case conceptualizations, it is possible for practitioners to show high rates of agreement on many aspects of a conceptualization, particularly those requiring fewer theory-driven inferences. Across other similar research designs, a consensus appears to have emerged that reliability in CBT case conceptualizations is greater at the level of descriptive information (symptoms and problems) and poorer at the level of more inferential information (cognitive or behavioural mechanisms in the maintenance of a person's difficulties), and which factors are most relevant within a particular case (Dudley, Park, James, & Dodgson, 2010; Kuyken, Padesky, & Dudley, 2009; Mumma & Smith, 2001; Persons & Bertagnolli, 1999).

Interestingly, in Kuyken and colleagues' discussion of their results (2005) the point is raised that given two therapists with different therapeutic orientations (e.g. behavioural and CBT) two quite different conceptualizations may be produced and yet both in some sense may be

"valid" or useful. They suggest that a greater expectation for reliability be made when therapists share the same major therapeutic orientation. They further speculate that any skillfully and collaboratively developed conceptualization might serve to help guide therapy to positive outcomes. Ultimately, the body of evidence available from research into CBT conceptualizations seems to indicate that they may be formulated across clinicians with at least a moderate level of reliability, particularly amongst more experienced clinicians and when utilizing structured formulation approaches (Kuyken, Padesky, & Dudley, 2009).

Following from indications that therapists show moderate levels of agreement on conceptualizations, it is reasonable to ask what the typical content and quality of a conceptualization may be. Eells, Kendjelic, and Lucas (1998) developed a comprehensive and theory driven system for evaluating case conceptualization content and quality. Their *Case Formulation Content Coding Method* (CFCCM) is a multi-theoretical system that built upon previous literature outlining the typical content that may be expected within case conceptualizations. The CFCCM has both content categories, relating to degree to which various possible types of information are present within a report, and several quality ratings. Quality ratings are given to each of the four main content categories: symptoms and problems, precipitating stressors, predisposing life events, and the inferred mechanism for linking the previous three categories and explaining a client's current difficulties. Additionally, ratings are made on the overall quality of the conceptualization in several areas; complexity of the conceptualization, how inferential the conceptualization is versus being merely descriptive, and how precise and tailored the language is within the conceptualization.

Results from their initial study utilizing the CFCCM in which they examined conceptualizations found in intake reports suggest that clinicians tend to use formulations

primarily to summarize the descriptive information regarding clients. Conceptualizations were evaluated, on average, as being relatively simple, only minimally inferential, and in many cases, lacking adequate information about potential mechanisms explaining the development and maintenance of a client's presenting issues. Later studies utilizing an expanded version of the CFCCM revealed that expert therapists produced higher quality conceptualizations across several domains, such as level of comprehensiveness, quality of inferred mechanisms, and goodness of fit to treatment plan (Eells, Lombart, Kendjelic, Turner, & Lucas, 2005).

Additional findings regarding the quality of case conceptualizations can be found in the previously discussed study by Kuyken et al. (2005), in which CCDs generated by workshop attendees were generated following a case presentation. In addition to examining the reliability of these conceptualizations, judgements were also made on their quality. Strikingly, according to their metric of quality only 44% of participants generated conceptualizations that were categorized as "good enough" or higher. Additionally a positive relationship (Spearman's rho = .22, $p < .05$, $N = 113$) between conceptualization quality and therapist experience (years of post qualification experience) was found. Additionally a chi-squared analysis suggested that the proportion of "good enough" conceptualizations improved incrementally across pre-qualified, qualified but non-BABCP (British Association for Behavioural and Cognitive Psychotherapies) accredited, and qualified and BABCP accredited clinicians.

Research conducted by Haarhoff, Flett, & Gibson (2011), using both the CFCCM and a similar measure of quality to that described in the previous study, found that few therapists included either biological or socio-cultural mechanisms within their conceptualizations. Therapists typically noted neither therapy interfering nor therapy enhancing factors and tended not to focus on the therapeutic alliance or protective factors within a client's life. They concluded

that 50-61% of therapist participants produced conceptualizations that were categorized as "good enough", a rating applied according to their study measure.

Despite the somewhat underwhelming results regarding the content and quality of case conceptualizations, one promising finding is that, given appropriate training, it appears this skill can be developed. Kendjelic and Eells (2007) examined the impact of a two hour training session on case conceptualization quality utilizing the CFCCM. Compared to those who did not receive the brief training, those in the training group generated more comprehensive conceptualizations that more often included elements from the four major categories of the CFCCM: symptoms and problems, precipitating stressors, predisposing life events, and an inferred mechanism. Across several quality categories the training group's conceptualizations were also superior. Global ratings of the conceptualizations indicated they tended to be more complex, more inferential, and had a greater precision in language. One major benefit articulated by the authors was that those in the training group began to use the case conceptualization as a tool for making inferences regarding possible mechanisms to explain the client's symptoms and problems, whereas participants in the control group primarily included descriptive information and were unlikely to put forward even rudimentary inferred mechanisms.

Although there may be some evidence that training can improve therapist generated conceptualizations, this issue would ultimately be of little relevance if case conceptualizations did not in some way provide a benefit to therapist or client in therapy. A preliminary issue then becomes whether conceptualizations actually substantially impact the direction of psychotherapy and the choice of interventions. Two studies provide some information on this issue. Dudley, Ingham, Sowerby, and Freeston (2015) tested whether case conceptualizations guide therapists to implement appropriate treatment strategies as well as whether level of therapist experience plays

a role in appropriate treatment selection. They found that when a comprehensive formulation is provided, therapists can generally select appropriate treatment options regardless of their level of experience. When required to build their own conceptualizations, expert therapists' conceptualizations were more parsimonious, internally consistent, and resulted in more appropriate ratings for treatment options when compared to the conceptualizations and judgements of less experienced therapists. These results suggest that given a well designed conceptualization, therapists may indeed be efficient and consistent in selecting treatment options for clients.

Other research (Groenier, Pieters, Witteman, & Lehmann, 2014) suggests a more nuanced and possibly problematic relationship between case conceptualizations and treatment decisions. The authors hypothesized that case conceptualization quality would positively relate to the complexity of a case, given the emphasis within the literature that case conceptualizations should find their greatest usefulness within the context of complex cases. They also hypothesized that the proposed mechanisms within conceptualizations should most strongly relate to treatment decisions, above other factors such as clinician background and orientation or DSM-IV classifications. This expectation stems from the notion that the proposed mechanisms in a case conceptualization should have the most treatment utility as they can be linked most directly to the mechanisms of change underlying specific interventions.

Interestingly, the authors determined that the reverse appeared to be true. More complicated cases were found to be associated with fewer causal factors being proposed, the incorporation of less relevant information, and overall lower conceptualization quality. Additionally the treatment decisions identified by clinicians were not linked to the proposed explanatory mechanisms in either high complexity or low complexity cases. Apparently the

treatment decisions did not stem from conceptualizations any more than from a clinician's background or the DSM-IV diagnoses clinicians ascribed to cases. The authors concluded that instead of utilizing case conceptualizations to guide treatment decisions, it appears that clinicians will more likely rely on guidelines or the presence of an available EST (in this case CBT) when presented with low complexity cases. For more complex cases, where an EST is not available, it appears that treatment decisions remain unconnected to the conceptualization and also exhibit greater variability in the specific treatments that are selected. Although methodological limitations (using case presentations, self selecting participants) of this study limit the strength of conclusions that can be drawn, it may be an important consideration that the very place where case conceptualizations should find their greatest use may be where they are least effectively applied (Groenier, Pieters, Witteman, & Lehmann, 2014).

If case conceptualizations do indeed guide therapy, and higher quality conceptualizations may more effectively guide therapists to selecting the appropriate treatments, the question of how case conceptualizations impact important therapy outcomes remains. In an early study on differences between manualized versus clinically flexible interventions within marital therapy (Jacobson et al., 1989), 30 couples were randomly assigned to one of two conditions: structured/manualized therapy or a clinically flexible condition. Following treatment, both conditions showed equivalent gains; however, the clinically flexible condition showed a superior retention of gains at a six month follow-up. These findings suggested some benefit of a more individualized treatment, in the form of a longer retention of treatment gains. Within the realm of the psychodynamic conceptualizations, Crits-Christoph, Cooper, and Luborsky (1988) demonstrated that the accuracy of a CCRT based interpretation predicted patient improvements during brief psychodynamic psychotherapy.

In a more recent study by Persons, Roberts, Zalecki, and Brechwald (2006), case conceptualization guided individualized CBT for anxious-depressed clients resulted in treatment gains comparable to those commonly obtained in RCTs for ESTs for either depression or anxiety disorders. Several additional studies comparing standardized versus individualized interventions have demonstrated some positive effects of individualization, such as a lower rate of non-responders to treatment for bulimia nervosa (Ghaderi, 2006). However, Kuyken, Padesky, & Dudley, (2009) note that the benefits for individualized and case conceptualization guided interventions tend to be small and limited to only select outcome measures.

Other research has either conflicted with the premise that case conceptualizations can enhance treatment outcomes, or at least not supported the superiority of case conceptualization guided interventions. In one study comparing tailor-made interventions and standardized therapy for phobic patients, the standardized condition was superior (Schulte, Künzel, Pepping, & Shulte-Bahrenberg, 1992). Chadwick, Williams, and Mackenzie (2003) examined whether developing and sharing formulations within CBT for drug-resistant patients with psychosis would impact client distress, symptoms, or the therapeutic alliance. Results of their two studies indicated that although therapists appeared to see the process of sharing the conceptualization as benefiting the alliance, and for some patients a rise in understanding and optimism occurred, no impact on symptoms was seen. The process of creating and sharing a case conceptualization did not in itself produce a direct impact on delusions, self-evaluations, or distress. Of note, it was found that for some patients the sharing of a conceptualization was both a positive and negative (i.e. mixed) experience, while other patients found it solely negative.

Recent research has again utilized the CFCCM to evaluate conceptualization quality and possible links between quality and therapy outcomes for individuals experiencing obsessive

compulsive disorder (Nattrass, Kellett, Hardy, & Ricketts, 2014). In this study, the sharing of a case conceptualization appeared to positively impact the alliance and reduce distress. However, case conceptualization quality as judged by the CFCCM was not related to outcome at any stage of treatment. This is similar to research conducted previously on the impact of individualized versus standardized treatments for OCD which showed equivalent gains across both interventions (Emmelkamp, Bouman, & Blaauw, 1994).

In 2009, Kuyken, Padesky, and Dudley articulated a new approach to case conceptualization with the publication of *Collaborative Case Conceptualization: Working Effectively with Clients in Cognitive-Behavioural Therapy*. The collaborative case conceptualization method (also described within Kuyken, Padesky, & Dudley, 2008) offered to explain some of the inconsistent results obtained in previous research and provides a new direction for the field to better test the links between case conceptualizations and treatment outcome. The model proposes an emphasis on four main principles of CBT case conceptualizations that the authors feel are critical to producing high quality and effective case conceptualizations.

The first area is levels of conceptualization. The authors argue that in past research designs therapists were required to quickly produce a single final case conceptualization based on a large amount of information shared all at once (Chadwick, et al. 2003; Persons, Mooney, & Padesky, 1995). This may not accurately reflect how case conceptualizations are actually developed in naturalistic therapies, where a therapist can begin by exploring more surface information while gradually learning more from the client and arriving at inferences about deeper processes at work such as triggers and maintenance factors and even longer term predisposing or protective factors. As such, the level and depth of a conceptualization can change

over time. The authors suggest that, in some cases, progression through levels of conceptualization can be rushed by therapists who attempt to include information in their formulation at a level too deep for the stage of therapy.

The second area within this approach is collaboration, which is the presence of a reciprocal and productive interaction between therapist and client. The therapist adds their expertise and knowledge of general psychological principles and models for psychopathology while listening carefully and collaborating with the client to create a mutually agreed upon conceptualization. In past research, the authors argue, conceptualizations have been somewhat one sided, and this may result in lower engagement of the client, fewer chances to refine and revise conceptualizations, ultimately reducing the rigour and usefulness of the case conceptualization.

In the third principle, empiricism, therapists ideally draw upon established nomothetic findings and models for psychopathology while actively testing their ideographic hypotheses about the client's difficulties. As a result of this empirical focus, therapists should receive rapid feedback for when their approach is incorrect or having poor results over the course of therapy. Overall the empirical approach should allow for a more adaptive therapy that the authors also suggest may have been missing in previous research.

Finally, the strengths and resiliency of clients are to be included in conceptualizations within the collaborative case conceptualization approach. Current CBT treatments may excessively focus on the negative particulars of a client's situation and this focus on problems, vulnerabilities, and adversity may limit the hopefulness and engagement of clients. By including elements of a client's resilience and strengths in the conceptualization guiding therapy, not only

can new avenues for reaching a client's goals be explored, but clients may feel less distress, and feel more empowered in therapy.

These principles highlight some of the proposed ways that case conceptualizations may interact with other therapy factors such as the therapeutic alliance, in that engagement, collaboration, and feeling that the therapy is on track may be improved by focusing on the four principles outlined above. As such, research exploring collaborative case conceptualization elements within CBT therapy may provide a more rigorous test for the usefulness of case conceptualizations than has been found in previous therapy guided by other conceptualization paradigms. To this end a manual, the Collaborative Case Conceptualization Rating Scale and Coding Manual (CCC-RS; Padesky, Kuyken, & Dudley, 2011) for coding and rating elements of the collaborative case conceptualization approach was developed and made available online. The CCC-RS contains 14 items roughly split amongst the four sub-scales/main principles of the approach. Each item is accompanied by detailed information on how to score therapists. Item scores range from a low of 0 to a high score of 3. This manual was developed to aid coders in scoring therapists on their ability to deploy the four main principles of collaborative case conceptualizations in live or pre-recorded therapy sessions.

Preliminary psychometrics of the CCC-RS have been recently published (Kuyken et al., 2015). Nine therapists with an average of 7.4 years experience had 40 of their audio recorded sessions evaluated by the study team utilizing the CCC-RS. The internal consistency and inter-rater reliabilities of the CCC-RS were both high, and total scale scores demonstrated a moderate correlation ($r = .54$, $p < .01$) with The Cognitive Therapy Rating Scale Revised (CTS-R; Blackburn et al., 2001), a scale developed to assess general CBT competence. The average overall score for the CCC-RS ($M = 18.90$, $SD = 7.84$) indicated an average item score of 1.4,

between the "novice" and "competent" levels. They found that client strengths were generally not a major focus in sessions and that no therapist's conceptualization demonstrated an "expert" level reflection of the most appropriate evidence based therapies (a score of three for that item). Instead they found most therapists generated conceptualizations based on generic CBT models. The authors concluded that the CCC-RS is a reliable measure for an important construct that demonstrates convergent validity with other measures of CBT skill. This suggests its use as a possible tool in the future of collaborative case conceptualization research.

There are a number of limitations to the extant body of case conceptualization research. These have been well articulated in past literature (Kuyken, Padesky, & Dudley, 2008) but will be described again here. First, the field has generally relied upon vignettes to elicit case conceptualizations from therapists. While a few noteworthy studies deviated from this pattern, the results regarding the quality and reliability of case conceptualizations from the remaining studies may not accurately represent case conceptualizations within therapy as usual (i.e., with real clients). Second, the findings regarding the link between case conceptualizations and therapy outcomes also must be interpreted cautiously. Across many of these studies, patients were selected based on a specific clinical disorder which may not represent the most useful place to deploy case conceptualizations and which also may limit the generalizability of findings. Third, in the studies comparing structured and unstructured interventions, evaluations of the quality of conceptualizations were also frequently absent. This makes interpreting the generally weak or negative associations between case conceptualization guided interventions and treatment outcome very difficult. It may be that across these studies the qualities of case conceptualizations were too poor to meaningfully improve therapy beyond the protocols developed for ESTs.

Finally, the definition for what constitutes quality within a conceptualization remains somewhat unclear. Within the CFCCM framework, content as well as the inferential, integrative and explanatory structure of the conceptualizations are considered markers of quality. Within the newly developed CCC-RS, quality appears to stem equally from the content of the conceptualization and the process through which it is developed. Given this recent shift in definition, further research may help to explore the links between case conceptualizations and outcome with a more rigorous approach than has yet been available.

Study 1: Adapting Two Case Conceptualization Coding Methods: Analysis of Inter-rater and Internal Reliabilities

Introduction

The process of psychotherapy involves the exchange of a great deal of information between client and therapist. This necessitates some process of synthesizing and distilling the information gathered in therapy sessions into a form that is more accessible to clinicians and can guide the direction of therapy. This process of integrating the experiences, background, symptoms, and goals of clients involves the development of a case conceptualization (also known as a case formulation). Case conceptualizations are seen as a key component of effective CBT by many authors (Persons, 2005; Beck, 1995; Needleman, 1999). Some arguments for the central role of case conceptualizations emphasize that this clinical tool has many useful benefits for therapy such as helping to promote insight and engagement in the client, helping to focus and prioritize which interventions to deploy, and validating and normalizing client experiences (Eells, 2011). In addition, case conceptualizations in psychotherapy can be seen as an alternative to what some clinicians perceive as a problematic trend towards more diagnosis-guided, rigid, and manualized approaches to psychotherapy (Restifo, 2011).

The main body of research on case conceptualizations ranges from studies on the reliability of case conceptualizations between therapists (Crits-Christoph, Cooper, & Luborsky, 1988; Persons, Mooney, & Padesky, 1995), to how brief training of case conceptualization techniques may increase case conceptualization quality (Kendjelic & Eells, 2007), how clinician experience relates to case conceptualization quality and content (Eells, Lombart, Kendjelic, Turner, & Lucas, 2005; Eells, et al. 2011), and how case conceptualizations guide treatment decisions among expert and novice clinicians (Dudley, Ingham, Sowerby, & Freeston, 2015).

A theme across many of these studies is that the quality of case conceptualizations developed by therapists generally appears to be somewhat lower than ideal, and that at more surface levels, such as the symptoms or diagnosis of a client, the reliability of conceptualizations across clinicians is greater than at deeper and more inferential levels, such as what underlying assumptions or beliefs may be driving a client's difficulties (Kuyken, Fothergill, Musa, & Chadwick, 2005). Within many of these research projects, structured methods for accessing case conceptualization quality have been developed and deployed.

The Case Formulation Content Coding Method (CFCCM) is one of the more widely utilized tools for evaluating conceptualizations in recent research designs. The appeal of the CFCCM may be that it was designed with the goal of being applicable across many psychotherapy orientations and that it examines both the content and quality of conceptualizations. Also, the authors drew upon several extant methods for constructing case conceptualizations and from the broader literature in order to direct the elements coded within the CFCCM. In the original paper describing its development (Eells, Kendjelic, & Lucas, 1998), the CFCCM is reported to have excellent psychometric properties with a mean inter-rater Kappa coefficient of .86 across the content and quality items of the measure, with coefficients ranging from .67 to 1.0. Within subsequent research, the CFCCM continued to demonstrate good reliability, and CFCCM scores were found to be higher for experienced therapists' conceptualizations when compared to those produced by novice therapists (Eells, Lombart, Kendjelic, Turner, & Lucas, 2005), suggesting some validity to the coding method. Recent research has applied the CFCCM in order to examine relationships between case conceptualization quality and therapy outcomes for individuals experiencing obsessive

compulsive disorder (Nattrass, Kellett, Hardy, & Ricketts, 2014); however, little relationship was found between quality and outcome.

In 2009, Kuyken, Padesky, and Dudley articulated a new approach to case conceptualization, the collaborative case conceptualization method, that offered to explain some of the inconsistent results obtained in previous research and offer a new direction for the field to better test the links between case conceptualizations and treatment outcome. The model proposed an emphasis on four main principles of CBT case conceptualizations that the authors felt were critical to producing high quality and effective case conceptualizations. The first area is levels of conceptualization, which proposes that early in therapy more surface information and descriptive conceptualizations may be appropriate, while later in therapy deeper processes at work, such as triggers and maintenance factors or predisposing or protective factors, may be integrated into the case formulation.

The second area within this approach is collaboration, which is the presence of a reciprocal and productive interaction between therapist and client. The therapist adds their expertise and knowledge of general psychological principles and models for psychopathology while listening carefully and collaborating with the client to create a mutually agreed upon conceptualization.

In the third principle, empiricism, therapists ideally draw upon established nomothetic findings and models for psychopathology, while actively testing their ideographic hypotheses about the client's difficulties through behavioural experiments, for example. As a result of this empirical focus therapists should receive rapid feedback for when their approach is incorrect or having poor results over the course of therapy.

Finally, the strengths and resiliency of clients are to be included in conceptualizations within the collaborative case conceptualization approach. Current CBT treatments may excessively focus on the negative particulars of a client's situation and this focus on problems, vulnerabilities, and adversity may limit the hopefulness and engagement of clients. By including elements of a client's resilience and strengths in the conceptualization guiding therapy, not only can new avenues for reaching a client's goals be explored, but clients may feel less distress, and feel more empowered in therapy.

A manual for coding and rating elements of the collaborative case conceptualization approach was developed and made available online, *the Collaborative Case Conceptualization Rating Scale and Coding Manual* (CCC-RS; Padesky, Kuyken, & Dudley, 2011). This manual was developed to aid coders in scoring therapists on their ability to deploy the four main principles of collaborative case conceptualizations in live or pre-recorded therapy sessions. In a recently published article, initial results on the psychometrics and reliabilities of the CCC-RS were reported (Kuyken et al. 2015). The scale demonstrated excellent internal consistency and inter-rater reliability and correlated moderately with the Cognitive Therapy Rating Scale (CTS-R; Blackburn et al., 2001).

Although the CCC-RS represents a viable and promising new avenue for exploring case conceptualizations developed in psychotherapy interventions, in its unaltered form it is designed to assess live or recorded therapy sessions. As demonstrated by the research stemming from the CFCCM (Eells et al., 1998), evaluating psychological/psychiatric reports for case conceptualization quality can produce interesting and informative results. As such, a version of the CCC-RS which can code for collaborative case conceptualization elements in therapy reports could also add significantly to the literature on case conceptualizations. To this end, the

psychometrics of a version of the CCC-RS adapted to do just that will be examined and contrasted with the CFCCM, as part of a larger study examining case conceptualizations in psychotherapy and their impact on the therapeutic alliance and treatment outcome.

Study Aims

This study represents a supplementary analysis of data as part of a larger study exploring the impact of case conceptualizations on psychotherapy outcomes and the therapeutic alliance. The reliabilities of two methods of evaluating case conceptualizations will be estimated, and their internal consistencies and structures will be evaluated. Beyond quantitative analyses, attention will also be given to the process of coding and the strengths and weaknesses of these two methods as executed by the study's coding team.

Method

Design and Sample

The study sample was comprised of 46 closed adult psychotherapy cases from the University of Waterloo's clinical psychology training clinic (the Center for Mental Health Research). These closed cases represented all therapy cases that had closed within the last three years where appropriate consent for research participation had been obtained. The study was reviewed by a university review board. The clients whose cases were included within this study's sample (Male = 16, Female = 30; Mean age = 33.08, *SD* = 12.48, range: 18 - 59) came from both the general population surrounding the clinic, and from the population of the University of Waterloo's undergraduate and graduate students.

A variety of presenting concerns were present across cases and clients met criteria for a variety of primary DSM-IV and DSM-V diagnoses (17% a unipolar depressive disorder, 13% social anxiety disorder, 6 % panic disorder with agoraphobia, 17% Generalized anxiety disorder, 7% NOS, other specified, or unspecified, anxiety disorder, 15% Obsessive Compulsive Disorder, 7% a personality disorder, 4% a simple phobia, 4% adjustment disorder, 2% Primary Insomnia). Twenty-two clients had at least one comorbid diagnosis, while five had no formal diagnoses at all. Clients came from a variety of cultural backgrounds, education levels, and marital statuses. Potential clients at the CMHR are referred elsewhere for services based on a limited set of exclusion criteria, including significant suicidality, legal involvement, current substance abuse, and active psychotic disorders. This sample of cases can be considered representative of the broader population of clients seen at this clinic.

Intervention

The therapists in this study were clinical psychology students ranging from their third year of clinical training through to therapists at the clinic for placement in their final internship year. All students received weekly supervision from registered clinical psychologists. The majority (90%) of cases were approached primarily from a cognitive behavioural (CBT) orientation; however approximately one quarter of cases included secondary therapy modalities (such as interpersonal therapy, Dialectical Behavioural Therapy, mindfulness/self-compassion, or problem solving/supportive work). Four therapy cases were not primarily CBT in orientation, being one of each of the following: assertiveness training, psychodynamic therapy, Interpersonal therapy, or psycho-educational. Within this minority of cases, elements of CBT were still present, and so were left in the sample.

Coding Measures

Collaborative Case Conceptualization Rating Scale and Coding Manual (CCC-RS): Developed by Padesky, Dudley, and Kuyken (2011), this manual guides coders in the rating of therapists' deployment of the four main principles of the collaborative case conceptualization method in either recorded or live therapy sessions. The CCC-RS contains 14 items roughly split amongst the four sub-scales/main principles. Item scores range from a low of 0 to a high score of 3. Given that the CCC-RS originally coded therapy sessions, modifications were required to several items in order to better match with the content that can be obtained within therapy reports. Specifically, the CCC-RS manual gives examples of behaviours that can be observed within sessions (i.e., "the therapist expresses a high degree of curiosity, interest, and ***detailed*** questions…" pp. 17), which were adapted to reflect more overall trends that could be coded from reports ("the report evidences a high level of curiosity and dedication to understanding the client..."). A brief description of our CCC-RS items can be found in Appendix C.

Case Formulation Content Coding Method v.1 (CFCCM): Developed by Eells, Kendjelic, and Lucas (1998), this method for assessing case conceptualizations in reports was generated to be applicable across several psychotherapy orientations. Inter-rater reliability of the CFCCM was high, the mean Kappa coefficient across categories was .86, with a minimum score of .67 and a maximum of 1.00 (Eells, Kendjelic, & Lucas, 1998). Scores on the CFCCM have been shown to be highest amongst therapists with expertise in case formulations (Eells, Lombart, Kendjelic, Turner, & Lucas, 2005). The CFCCM v.1 has both content categories, relating to the degree various possible types of information are present within a report (0- absent, 1 -somewhat present, 2- clearly present), and several quality ratings. Quality ratings (from 1-4) are given to each of the four main content categories; symptoms and problems, precipitating stressors,

predisposing life events, and inferred mechanism for linking the previous three categories and explaining client's current difficulties. Additionally, ratings are made on the overall quality of the conceptualization in several areas: complexity of the conceptualization, how inferential the conceptualization is versus being merely descriptive, and how precise and tailored the language is within the conceptualization.

For the purposes of this study several changes were made to the CFCCM to simplify the coding, eliminate content categories that were predicted to have an extremely low base rates, be more comparable to scoring on the CCC-RS, and re-orient some of the coding to focus more on the quality and integration of the content categories as opposed to solely their degree of presence.

In this altered version each remaining content category received a score from 0 to 3. A score of 0 indicated that a particular category was absent. A score of 1 indicated the content was present in a very limited or ambiguous form without integration with other elements of the conceptualization. A score of 2 indicated at least one clear mention of this content category with adequate integration with the rest of the conceptualization in a way that may inform the direction of therapy. A score of 3 was given for outstanding examples of a content category such that it was clear how this information being included could guide therapy with good links to other elements of the conceptualization.

Several content categories of the CFCCM also contain sub-categories, such as the category of inferred mechanism which has subcategories for what type of mechanism (psychological, biological, socio-cultural, or substance abuse) is being proposed as the mechanism for the maintenance of the client's problems. In the study first outlining the CFCCM, these items were coded across three levels: not present, somewhat present, and clearly present.

However, the authors found disagreement between the somewhat present and clearly present levels and thus collapsed these items to either present or absent, a strategy which we incorporated into the scoring of our subcategories at the outset. Finally, similar to other research with the CFCCM (Nattrass, Kellett, Hardy, & Ricketts, 2014), an overall quality item was generated based on a coder's overall impression of the conceptualization and based on the previous elements of the CFCCM. For a summary of all items, subscales, and sub-categories found across the adapted CFCCM and the CCC-RS utilized in this study see Table C1 within Appendix C.

Procedure

1) Adapt and Prepare Coding Methods: As conceptualizations were coded in this study without recordings of therapy sessions, the collaborative case conceptualization rating scale and content coding method required modifications to be applicable and relevant to case reports while maintaining focus on the constructs of interest in the original coding schemes.

2) Select and prepare case reports: Files were accessed and any assessment, progress, or discharge reports within each case were copied with client names and birthdates, and therapist and supervisor names redacted.

3) Train team of coders: The team of coders consisted of two second year master's level students in clinical psychology and three undergraduate students in psychology. Training first involved introducing the undergraduate students to the broader structure and tools of CBT therapy (assessment and treatment planning, automatic thoughts, core beliefs, thought records, behavioral experiments, exposure therapy, etc.) as well as to other important concepts being coded in our study (such as the working alliance, therapy engagement, therapy dropout, and case

complexity). Second, coders met over several weeks to discuss the two coding methods and the additional coding items utilized in the study. Each item's description was reviewed together and items which remained unclear to any coder were identified and further information was added to the manual to guide coding. As well, examples of how each item might be represented in case reports were discussed. The final step in training was to meet over several weeks to code practice reports as a team until coders felt confident in their ability to independently code the sample of files. In addition to group training sessions, undergraduate coders were encouraged to contact the first author individually for more information on psychotherapy concepts and procedures if the content or concepts within any report was unclear.

4) Code the reports and assess reliabilities: Following training, the 46 sets of reports were then coded independently over the course of six weeks. Both assessment and discharge reports for each case received independent scores for each item of the CCC-RS and the CFCCM. Alliance ratings were made on the therapy case as a whole. Meetings were held weekly to discuss the general progress of coding, identify and discuss whether any particular items were presenting challenges, and to help ensure coder drift did not occur. Halfway through coding the corpus of reports, reliabilities for items were calculated and brought to the weekly meeting. Items with low reliability were reviewed and methods to improve consistency of ratings were identified.

5) Gather Feedback from Coders: Each coder was invited to submit a short reflection on their experience coding the sample of reports. In particular it was hoped that coders would be willing to articulate which items from the coding methods they found most challenging to code and why this may have been.

Data Analysis

Utilizing SPSS statistical software v.22, intraclass correlation coefficients (ICCs) (Shrout & Lane, 2012) were calculated for the numerical variables coded, and Fleiss' Kappa statistics (Fleiss, 1971) were calculated for nominal (i.e. present/absent) items of the CFCCM. Reliabilities were calculated twice for each item, once for the ratings made from the assessment reports and again from those made from the discharge reports. Due to the nature of the research method, no cases contained missing data across raters.

Results

Shrout (1998) proposed guidelines for interpreting measure reliability. Reliability values of .00 - .10 show "virtually no reliability"; .11 -.40 show "slight" reliability; .41 - .60, "fair" reliability; .61 - .80, "moderate"; and .81 - 1.00, "substantial" reliability. Negative values can indicate systematic disagreement between raters. In discussing desired ranges for various research purposes Shrout and Lane (2012) state a reliability of above .80 is desirable for more definitive research, but preliminary studies may be conducted with reliabilities of at least .50.

Reliability assessed through consistency ICCs for the collaborative case conceptualization items at assessment ranged from a low of -.07 for the item on parsimony, to a high of .73 for the item on interest in client strengths. The average ICC across all 14 items of the CCC-RS was.53, which falls in the "fair" range. The lowest ICC value for the scalar items of the CFCCM was .13, for the precision of language item. The highest was .68 for the inferred mechanism item. The average of the ICCs from the CFCCM at assessment was .54, again falling in the fair range for reliability. Similar, though slightly higher results were obtained for ICC values for the discharge report reliabilities. The breakdown of reliabilities for each scalar item of

the CCC-RS and CFCCM at assessment can be found in Table 1. Interested readers can consult Table A1 within Appendix A for the discharge reliabilities.

In order to determine the possible causes of the modest reliability across the two subscales, corrected item-total correlations for each coder were produced for each item. Within these analyses each coder was an "item" of the scale (item on the CCC-RS of CFCCM). Results of these analyses indicated that across many items reliability could be improved by dropping coders. Within 19 items from the assessment coding (70%), one or two undergraduate coders (not always the same) could be dropped resulting in an improved reliability. In only two items from the assessment coding was one of the graduate level coders showing the weakest item-total correlation. Additionally, six items appeared to be at their maximum reliability, such that coders could not be dropped to improve the ICC value. Again, similar results were found for the discharge items.

The absent/present subcategory items of the CFCCM showed overall poorer reliability than the scalar items of the CCC-RS and CFCCM. From the items representing the assessment reports Fleiss' Kappa values ranged from -.05 to .31, averaging .20 across all items (Fleiss' Kappa reliabilities for each item are reported in Table 2 for the assessment coding. Discharge reliabilities are found in Appendix A, Table A2). Reliabilities within this range may be interpreted as "slight" and below most acceptable cut-offs for research purposes. One consideration regarding these values is that when a majority of coders agree that an item is present or absent most of the time, then even infrequent false positives or negatives can result in low reliabilities as the proportion of error variance will be high in comparison to the actual variance across the two categories of the dichotomous variable. As such prevalence rates for the construct being coded for can heavily influence item reliabilities. This issue has caused some

authors to suggest that reliability analyses for dichotomous variables may not be appropriate or useful (for a discussion and further readings, see Shrout & Lane, 2012). Within several of the dichotomous CFCCM items coded in this study, the prevalence issue may have skewed reliability estimates downward despite significant agreement amongst raters.

To evaluate this possibility, the average percent absolute agreement of ratings across all pairs of raters was calculated for the dichotomous items. For six items, the average rates of absolute agreement across these pairs were high enough to suggest that despite a majority of raters agreeing on the presence or absence of an item, some false positives or negatives were dramatically reducing the Fleiss' Kappa reliability obtained. As an example of this, the lowest reliability item of the CFCCM dichotomous items was for the inferred mechanism subcategory "inferred psychological mechanism," and yet the average absolute agreement across each pair of raters was above 90%. Although tempting to explain the low average reliability for the dichotomous items as resulting entirely from these prevalence issues, other items showed both low reliability and low average absolute agreement across raters.

As described previously, reliabilities tended to increase upon removal of two undergraduate coders across a majority of items. Given the independence of this procedure to any outcome variable, it was decided that all item reliabilities would be recalculated after removing the two coders contributing least to the systematic variance within an item (Tables 1, 2, A1, and A2 can be referred to for a detailed breakdown of which coders were dropped per item). This process improved the average inter-rater reliability for the CCC-RS items, at assessment from a .53 to .60, and for the CFCCM scalar items, from a .54 to a .61. The dichotomous CFCCM items were also recalculated with the lowest coders dropped and although there was an improvement in average Fleiss' Kappa from .20 to .35 this still left these items

generally below acceptable cutoffs for use in research. Two additional items, item four of the CCC-RS, parsimony of the conceptualization, and the overall precision of language item within the CFCCM, item 12, also remained below acceptable cutoffs.

Following the examination of inter-rater reliabilities, attention was turned to the internal reliability of the various scale totals and subscales of both coding methods. First, an average score across raters was calculated for each scale item of each case, once for codes at assessment and again for codes at discharge. This process was repeated twice, once with all coders included, and once with the lowest reliability coder(s) dropped from each item. This allowed for the impact of dropping coders on item means and scale internal reliabilities to be examined. Means and standard deviations for each scalar item of the CCC-RS and CFCCM can be seen in Appendix C, Table C2 for assessment coding, and Table C3 for discharge.

Internal reliabilities were estimated through Cronbach's alpha statistics. Dropping of two coders appeared to have had a generally minimal impact on the internal reliabilities obtained and as such the remainder of the article will generally focus on values obtained from the items representing the highest reliability coders. The Levels of conceptualization subscale of the CCC-RS consists of items one to four, and produced a Cronbach's alpha of .85, demonstrating good internal consistency. The Collaboration subscale is comprised of items 5, 6, and 7; this subscale had very poor internal consistency, with an alpha value of .35. The Empiricism subscale, items eight through ten, had an alpha value of .64, a moderate value falling within the questionable range of consistencies. The Strengths and Resilience subscale, items 11 to 14, had good consistency with an alpha value of .80.

Two issues regarding the CCC-RS suggested that subscale items should be dropped. Given the poor inter-rater reliability of item 4, this item was dropped from the Levels subscale. Additionally, the low internal reliability for the Collaboration subscale appeared to stem from item 6, which upon further examination did not significantly correlate with either item 5 or item 7. As can be seen in Table C2, the mean score for this item was significantly lower than the other two items of the subscale. The item likely suffered from a restricted range and could not correlate with the other two items due to a floor effect. Cronbach's Alpha reliabilities were recalculated for the Levels and Collaboration subscales with items 4 and 6 dropped, respectively. While the Levels subscale value slightly fell, the Collaboration subscale dramatically increased from .35 to .60. However, a two item subscale may be somewhat limited in scope, and a reliability of .60 remains somewhat short of the ideal.

Utilizing the Spearman-Brown prophecy formula, an estimate can be produced for the number of items that would be required to increase a test's reliability to a desired value given the current average reliability of test items. To achieve a Cronbach's alpha value of .80 for the Levels subscale, an additional three to four items would be required. As well, the addition of four items to the Empiricism subscale would allow for a reliability of .80 to be reached. With the dropped items excluded, Cronbach's alpha for the CCC-RS scale total was .86, suggesting a good internal consistency across the whole scale.

For the CFCCM, distinct subscales are less apparent, and yet the original developers articulate a few possible ways items may be grouped together (Eells, et al., 1998). First, they suggest that items 1 through 4 are common elements that are found across a variety of case conceptualization paradigms. Second, the overall quality items are similar in that they are rated on the conceptualization as a whole and are not tied specifically to the content present. Within

these two sets of items, it is not clear that there is an underlying construct driving scores as in the CCC-RS subscales. As such, it is most appropriate to calculate a composite score for their reliabilities in lieu of a Cronbach's alpha value. A typical composite score utilizes the reliability of separate multi-item tests; however, for our purposes the calculations relied upon the reliability of each item being composited, as estimated by the inter-rater reliability obtained via the previously described ICC calculations. The composite reliabilities for these two item groupings were good; the reliability for the composite derived from items 1-4.was .81 at assessment and .84 at discharge. The reliability for the overall items was .81 at assessment and .83 at discharge (internal reliabilities at assessment are presented within Table 3, See Appendix A, Table A3 for discharge coding values).

Table 1. Inter-rater Reliabilities for CCC-RS and CFCCM Items from Assessment Coding

CCC-RS	ICC (all coders)	ICC (lowest dropped)	Coders Dropped
Levels of Conceptualization Items			
1. Clear Link to Goals	.60	.69	Und. 1, Und. 2
2. Clear Rationale and Engagement	.48	.52	Und. 2
3. Meaningful Account of Issues	.43	.46	Und. 2
4. Good Parsimony	-.07	.19	Und. 1, Und. 2
Collaboration Items			
5. Collaboratively Developed CC	.56	.64	Und. 1, Und. 2
6. Culture and Experience	.66	.74	Und. 2
7. Genuine Curiosity	.67	.71	Und. 1
Empiricism Items			
8. Justified CBT model	.65	.65	N/A
9. Test of "fit" of conceptualization	.41	.53	Und.1, Und.3
10. Treatment linked to CC	.60	.60	N/A
Strength and Resilience Focus Items			
11. Interest in Client Strengths	.73	.73	N/A
12. Client Strengths and Treatment	.60	.67	Und. 1, Und.3
13. Client Aspiration Focus	.54	.54	N/A
14. Client Resilience Focus	.60	.66	Und.1, Und.3
Average ICC all items	.53	.60	

CFCCM	ICC (all coders)	ICC (lowest dropped)	Coders Dropped
Content Quality Ratings			
1. Symptoms and Problems	.40	.53	Und 1., Grad. 1
2. Precipitating Stressors & Events	.69	.71	Und. 1
3. Predisposing Life Events	.62	.66	Und. 1
4. Inferred Mechanism	.68	.71	Und. 1
5. Client History Categories	.59	.59	N/A
6. Iatrogenic Factors	.41	.45	Grad. 2
7. Global Level of Adjustment	.58	.58	N/A
8. Treatment Indicators	.57	.59	Und. 3
9. Therapist T_x Expectations	.67	.70	Und. 2, Und. 3
Overall Quality Ratings			
10. Complexity of Formulation	.63	.71	Und. 1
11. Degree of Inference	.63	.67	Und. 1, Und. 3
12. Precision of Language	.13	.35	Und. 1, Und 2
13. Overall Formulation Quality	.37	.62	Und. 1, Und 2
Average ICC All items	.54	.61	

Table 2. Inter-rater Reliabilities for CFCCM Present/Absent items (Assessment Coding)

	Fleiss' Kappa (all coders)	Fleiss' Kappa (lowest dropped)	Coders Dropped
Inferred Mechanism			
a. Inferred Mech. Psychological	-.05	-.05	N/A
b. Inferred Mech. Biological	.33	.46	Und. 1, Und. 2
c. Inferred Mech. Socio-Cultural	.15	.36	Und. 3, Und. 1
d. Inferred Mech. Substance Use	.66	.66	N/A
Client History Categories			
a. Own or Family Psych. History	.29	.53	Und. 3, Und. 1
b. Own or Family Medical History	.22	.53	Grad. 1, Und. 1
c. Developmental or Social History	.11	.19	Und. 3, Und. 2
Treatment Indicators			
a. Negative Treatment Motivation	.23	.36	Und. 1, Und. 2
b. Positive Motivation for Treatment	.10	.32	Und. 3, Und. 1
c. Positive Social Support	.26	.30	Und. 3
d. Posit. self perception, goal, wish	.06	.30	Und. 3
Therapist Treatment Expectations			
a. Negative Treatment Indications	.31	.36	Und. 2
b. Prognosis	-.02	.17	Und. 2, Und. 1
Average Fleiss' Kappa	.20	.35	

Table 3. Internal Reliabilities for Coding Methods and Subscales (Assessment Coding)

	Cronbach's α	
	All Coders	Lowest Dropped
Levels of Conceptualization Subscale	.85	.85
Levels of Conceptualization (Item 4 dropped)	.81	.79
Collaboration Subscale	.39	.35
Collaboration Subscale (Item 6 dropped)	.58	.60
Empiricism Subscale	.71	.64
Strength and Resilience Subscale	.78	.80
CCC-RS Total Scale (all items)	.86	.86
CCC-RS Total (Items 4 and 6 dropped)	.86	.85
	Composite Scale Reliability	
CFCCM Items 1-4	.81	.81
CFCCM Overall Items	.76	.82

Discussion

The Collaborative Case Conceptualization Rating Scale

Inter-rater reliability for our modified version of the CCC-RS was moderate at the level of specific items. Although some items showed higher reliability across all five coders, such as the interest in strengths item, other items remained poor even after retaining only the most reliable raters, such as the parsimony item. At the level of its subscales, reliability for the CCC-RS appeared somewhat stronger, at least following the removal of item 6 from Collaboration. Internal consistency was strongest for the Levels of Conceptualization and Strengths and Resilience focus subscales, both of which originally contained four items each. Dropping of the unreliably coded item 4 somewhat reduced the internal consistency for the Levels subscale. Taken with the Collaboration and Empiricism subscales' somewhat poorer internal consistencies, it may be that additional items are necessary to improve CCC-RS subscale reliabilities, at least within the context of coding reports.

Although the original version of this rating method has not yet been utilized widely, a recently published article (Kuyken et al., 2015) has described its preliminary psychometrics. Several noteworthy similarities between their results and ours can be found. First, elements of a client's culture were found to be underutilized by clinicians across both studies. Indeed, within the present study, the incorporation of cultural considerations into conceptualizations was uncorrelated with the other two items from the collaboration subscale and this may warrant some attention to this item's inclusion in later research with the CCC-RS. Second, items 11 and 13 were found to have poor item-total correlations by Kuyken et al. (2015) and within our study we found the Strengths and Resilience subscale correlated less with other subscales in general. In fact it did not significantly correlate with any of the other subscales at assessment, and at

discharge it appeared to less strongly inter-correlate than the other subscales. While this does not speak to the usefulness or benefits of a strengths and resiliency focus in therapy, it does suggest that this component of the CCC-RS may represent a more distinct skill, mindset, or approach than the constructs guiding the remainder of the CCC-RS.

The reliabilities within this study were somewhat lower than those reported by Kuyken's group; however our study contained a much less homogeneous sample and utilized a team comprised of both undergraduate and graduate level coders. As well, the method of our study involved coding CCC-RS items from case reports and involved modifying the scale for this purpose. Some items may have be less accessible to coding through reports, either due to less clear markers for item quality or due to more ambiguous material being coded. Despite this, the CCC-RS items coded within this study did still fall above cut-offs for inclusion in research designs, with the exception of one item: a particular challenge appeared for our team of coders when evaluating the parsimony of conceptualizations from reports. Explanations from our coders suggest that it was difficult to determine what information truly was or was not relevant to include in a conceptualization, particularly with only one limited viewpoint into a particular case (i.e., that provided by the reports). Additionally, coders expressed that judgments could be more easily made at the extremes of problems with parsimony, where there was far too much or far too little information and the conceptualization appeared problematically limited or problematically complex.

The Case Formulation Content Coding Method

Reliability statistics for the CFCCM major categories evaluating quality and presence were generally greater than for the subcategories. Similar findings have apparently also been

found by Eells, et al., (2011), who state that the specific agreement on item subcategories was lower than for their overall category ratings. Despite this similarity, it remains that our reliabilities on these finer grained categories were markedly low, below most cut-offs for research purposes. As previously discussed, some of the low Fleiss' Kappa reliabilities may be the result of prevalence/base rate issues. For those subcategories dealing with socio-cultural content, the low reliabilities may have also stemmed from the content of the reports themselves. When therapists discussed the social sphere of clients, they often appeared to mix both internal factors and external factors, such as noting a lack of social support but also noting a client's social anxiety regarding going to see friends, for example. As such, some of our raters may have identified these types of statements more as psychological mechanisms, and other raters more as socio-cultural. These socio-cultural factors also appeared generally less elaborated upon within reports which may be reflected in the low mean of item 6 (cultural experience) of the CCC-RS. The low reliability for the overall degree of precision of the language item at assessment could be the result of a lack of clear markers of quality identified by our group for this item.

Overall Impressions

The high internal reliabilities obtained across most of the prescribed or derived subscales suggest that the most appropriate level of analysis for research purposes may not lie at the item level within these scales. Our findings also suggest that the addition of several items to the collaboration and empiricism subscales may improve the internal reliabilities of the CCC-RS when applied to report coding. Given the nature of the collaboration subscale construct it could be possible to adapt and include items from measures of the therapeutic alliance. Alternatively, it may also be possible to boost both inter-rater and internal consistency scores by utilizing a coding team with more experience in clinical settings. Although considerable time was spent

training our undergraduate coders, it appears as though an experience base may be important for accurate coding. Finally, it appears as though discharge report coding was generally more reliable than assessment coding. One possible explanation for this is that discharge reports were often shorter and more focused in the material presented. This in turn may have taxed coders less, or led to more agreement in ratings due to greater consistency in discharge than in assessment reports. Additionally, discharge conceptualizations may be more reliably coded as they represent more well developed conceptualizations from the end of therapy, as opposed to the initial conceptualizations developed at assessment.

Study 2: Case Conceptualizations in Psychotherapy Reports: Relationships to Outcome and the Alliance

Introduction

The process of psychotherapy involves the exchange of a great deal of information between client and therapist. This necessitates some process of synthesizing and distilling the information gathered in therapy sessions into a form that is more accessible to clinicians and can guide the direction of therapy. This process of integrating the experiences, background, symptoms, and goals of clients represents the development of a case conceptualization (also known as a case formulation). Case conceptualizations are seen as a key component of effective cognitive behavioural therapy by many authors (Beck, 1995; Needleman, 1999; Persons, 2005). Some arguments for the central role of case conceptualizations emphasize that this clinical tool has many useful benefits for therapy such as helping to promote client insight and engagement, helping to focus and prioritize which interventions to deploy, and validating and normalizing client experiences (Eells, 2011). In addition, case conceptualizations in psychotherapy can be seen as an alternative to what some clinicians perceive as a problematic trend towards more diagnosis-guided, rigid, and manualized approaches to psychotherapy (Restifo, 2011). These manualized/structured approaches have also been seen as too limited to effectively treat clients whose difficulties are more complicated or intense than those found in the randomized control trials (RCTs) in which manualized treatments are often developed (Persons, 2005).

One difficulty that has kept the debate alive between using idiographic, client-tailored case conceptualizations versus deploying more manualized treatments guided by nomothetic models of psychopathologies is the paucity of research into the use of case conceptualizations in psychotherapy. This is particularly true regarding investigations of the practical effects of case

conceptualizations on therapy outcome and related therapy variables (Bieling & Kuyken, 2003). Other areas in case conceptualization research have received relatively more attention. Research has been conducted to examine the reliability of case conceptualizations between therapists (Crits-Christoph et al. 1988; Persons, Mooney, & Padesky, 1995), how brief training of case conceptualization techniques may increase case conceptualization quality (Kendjelic & Eells, 2007), how clinician experience relates to case conceptualization quality and content (Eells, Lombart, Kendjelic, Turner, & Lucas, 2005; Eells, et al. 2011), and how case conceptualizations guide treatment decisions among expert and novice clinicians (Dudley, Ingham, Sowerby, & Freeston, 2015). A theme across many of these studies is that the quality of case conceptualizations developed by therapists generally appears to be somewhat lower than ideal, and that at more surface levels, such as the symptoms or diagnosis of a client, the reliability of conceptualizations across clinicians is greater than at deeper and more inferential levels, such as what underlying assumptions or beliefs may be driving a client's difficulties (Kuyken, Fothergill, Musa, Chadwick, 2005).

The limited research that has been conducted examining the relation between case conceptualization quality and outcome has shown mixed results. In an early study on differences between manualized versus clinically flexible interventions within marital therapy (Jacobson et al., 1989), 30 couples were randomly assigned to one of two conditions: structured/manualized therapy or a clinically flexible condition. Following treatment both conditions showed equivalent gains; however, the clinically flexible condition showed a superior retention of gains at a six month follow-up. This suggested some benefit of a more individualized treatment, and therefore an ideographic conceptualization approach, in the form of a longer retention of treatment gains.

In a study by Persons, Roberts, Zalecki, and Brechwald (2006), treatment gains were deemed to be comparable between case conceptualization guided CBT for anxious-depressed clients in comparison to results commonly obtained in RCTs for manualized or empirically supported therapies (ESTs). This finding suggested that for patients with comorbidities, case conceptualization guided therapy may represent a viable and empirically supported option.

Other research has either conflicted with these promising results, or at least not supported the superiority of case conceptualization guided interventions. One study comparing tailor-made interventions and standardized therapy for phobic patients indicated that the standardized condition was superior (Schulte, Künzel, Pepping, & Shulte-Bahrenberg, 1992). Chadwick, Williams, and Mackenzie (2003) conducted two experiments to investigate whether developing and sharing formulations within CBT for drug-resistant psychosis patients would impact client distress, symptoms, or the therapeutic alliance. Results of these two studies indicated that although therapists appeared to see the process of sharing the conceptualization as benefiting the alliance, and for some patients a rise in understanding and optimism occurred, no impact on symptoms was observed. The process of creating and sharing a case conceptualization did not in itself produce a direct impact on delusions, self-evaluations, or distress. Of note, it was also found that the experience was both a positive and negative (i.e. mixed), or solely negative, experience for some patients.

The Collaborative Case Conceptualization Method

In 2009, Kuyken, Padesky, and Dudley articulated a new approach to case conceptualization, the collaborative case conceptualization method, that offered to explain some of the inconsistent results obtained in previous research and offer a new direction for the field to

better test the links between case conceptualizations and treatment outcome. The model proposed an emphasis on four main principles of CBT case conceptualizations that the authors felt critical to producing high quality and effective case conceptualizations. The authors refer to their first area as "levels of conceptualization." The authors argue that in past research designs, therapists were required to quickly produce a single final case conceptualization based on a large amount of information shared all at once (Chadwick, et al. 2003; Persons et al. 1995). This may not accurately reflect how case conceptualizations are actually developed in naturalistic therapies, where a therapist can begin by exploring more surface information while gradually learning more from the client and arriving at deeper process at work such as triggers and maintenance factors and even long term predisposing or protective factors. As such, the level and depth of a conceptualization can change over time and may be rushed by therapists who attempt to include information in their formulation at a level too deep for the stage of therapy.

The second area within this approach is collaboration, which is the presence of a reciprocal and productive interaction between therapist and client. The therapist contributes their expertise and knowledge of general psychological principles and models for psychopathology while listening carefully and collaborating with the client to create a mutually agreed upon conceptualization of the client's difficulties. The authors argue that in past research conceptualizations have been somewhat one sided, resulting in lower engagement of the client, fewer chances to refine and revise conceptualizations, and consequently a reduction in the rigour and usefulness of the case conceptualization.

The third principle, empiricism, refers to therapists drawing upon established nomothetic findings and models for psychopathology, while actively testing their idiographic hypotheses about the client's difficulties. As a result of this empirical focus, therapists should receive rapid

feedback for when their approach is incorrect or having poor results over the course of therapy. Overall the empirical approach should allow for a more adaptive therapy that the authors also suggest may have been missing in previous research.

Finally, the authors argue that the strengths and resiliency of clients ought to be included in collaborate case conceptualizations. Current CBT treatments may excessively focus on the negative particulars of a client's situation and this focus on problems, vulnerabilities, and adversity may limit the hopefulness and engagement of clients. By including elements of a client's resilience and strengths in the conceptualization guiding therapy, not only can new avenues for reaching a client's goals be explored, but clients may feel less distress, and feel more empowered as they engage in the process of change.

These principles highlight some of the proposed ways that case conceptualizations may interact with other therapy factors such as the therapeutic alliance, engagement, a sense of collaboration, and feeling that the therapy is on track may be improved by focusing on the four principles outlined above. As such, research exploring collaborative case conceptualization elements within CBT therapy may provide a more rigorous test for the usefulness of case conceptualizations than has been found previously. To this end, a manual for coding and rating elements of the collaborative case conceptualization approach was developed and made available online: *the Collaborative Case Conceptualization Rating Scale and Coding Manual* (CCC-RS; Padesky, Kuyken, & Dudley, 2011). This manual was developed to aid coders in scoring therapists on their ability to deploy the four main principles of collaborative case conceptualizations in live or pre-recorded therapy sessions.

The CCC-RS suggests a new approach to evaluating case conceptualizations that focuses on both the content and process of conceptualizations within CBT. By evaluating process elements such as whether the level of a conceptualization appears appropriate for the stage of treatment or whether both client and therapist are contributing ideas and listening carefully to the other the CCC-RS adds new tools absent from previous research that often focused on conceptualizations generated unilaterally and without the possibility for improvements and growth over time. The CCC-RS also emphasizes the importance of specific content within conceptualizations. Each main principle also includes specific content that can be leveraged to improve therapy such as client goals and aspirations, an empirically based model, and evidence of prior resilience.

The Case Formulation Content Coding Method

Developed by Eells, Kendjelic, and Lucas (1998), the Case Formulation Content Coding Method (CFCCM) is listed as one of the case conceptualization evaluation tools that informed the development of the CCC-RS, and yet the two differ in terms of primary focus. Primarily, the CFCCM can be utilized for categorizing and evaluating, as the name suggests, the content of case conceptualizations with little focus on the process by which a conceptualization is developed. Secondarily, it provides some ratings of quality relating to the integration and elaboration of the content within certain categories. Advantageously, the content which the CFCCM codes for is general enough to span across many psychotherapeutic orientations yet emphasizes important categories, outlined later, found across the majority of case conceptualization approaches.

The CFCCM can be reliably applied to evaluate case conceptualizations found in psychological/psychiatric reports, and has also found use in evaluating conceptualizations within session transcripts and other sources (Eells, Lombart, Kendjelic, Turner, & Lucas, 2005). Recent research has also explored the relationship between CFCCM ratings and therapy outcomes for individuals experiencing obsessive compulsive disorder (Nattrass, Kellett, Hardy, & Ricketts, 2014). In this study the sharing of a case conceptualization did positively impact the alliance and reduce distress; however, case conceptualization quality as judged by the CFCCM was not related to outcome at any stage of treatment. Given the comprehensive and adaptable nature of the CFCCM and its limited use evaluating case conceptualization outcome links to date, it remains a viable tool for further case conceptualization research.

Current Study Aims

While previous research designs have evaluated case conceptualizations found within written reports, the current study expands upon this approach by additionally evaluating possible links between case conceptualizations and treatment outcome. Additionally, this research includes two methods for evaluating case conceptualizations: the CFCCM and the CCC-RS, which represent two complimentary and contrasting methods for approaching case conceptualization research.

The CFCCM has shown good reliability in previous research designs and, through its focus, may allow judgements to be made on how case conceptualization content relates to therapy outcome. While some research on this has been conducted showing little relation, (Nattrass et al., 2014), such work examined case conceptualizations within a more homogeneous sample of OCD patients and may not generalize widely. We predict that the quality of the four

major content items of the CFCCM, the range of content in the conceptualization, as well as the overall quality of conceptualization as judged by the CFCCM will predict therapy outcomes.

As articulated previously, the CCC-RS and the collaborative case conceptualization method may have several possible advantages when compared to previous methods. Related to this, it represents a focus on both the content and process of a conceptualization, which may be more important factors to consider than the content and structure of a conceptualization alone (Nattrass, 2014). Within this study, ratings of case conceptualization quality made utilizing a modified version of the CCC-RS (altered to be more applicable to case reports) will be evaluated in relation to treatment outcome and the alliance. A positive relationship is expected where higher ratings of the CCC-RS will predict both greater treatment outcomes and a higher therapeutic alliance. Additionally, a greater relationship may be seen between the CCC-RS and alliance than for the CFCCM, given the emphasis on collaboration and overall process contained in the CCC-RS.

Method

Design and Sample

Forty-six closed adult psychotherapy case files from the archives of the Centre for Mental Health Research (the University of Waterloo's clinical psychology training clinic) were selected for use in the present study. These represented all adult therapy cases that had closed within the last three years where appropriate consent for research participation had been obtained. Files contained diagnostic and symptom measures, assessment reports, and discharge reports. Moreover, for some longer therapy cases, a mid-treatment progress report is also written and available for analysis in the present study. Two files within the sample only contained

assessment reports - one due to client drop out after the assessment phase, the other because a client was referred out of the clinic after the assessment was completed; these cases were included in the study sample.

Clients (Male = 16, Female = 30; Mean age = 33.08, *SD* = 12.48, range: 18 - 59) came from both the general population surrounding the clinic, and from the population of University of Waterloo undergraduate and graduate students. A variety of presenting concerns were present across cases and clients met criteria for a variety of primary DSM-IV and DSM-V diagnoses, including17% with unipolar depressive disorder, 13% social anxiety disorder, 6 % panic disorder with agoraphobia, 17% Generalized anxiety disorder, 7% NOS, other specified, or unspecified, anxiety disorder, 15% Obsessive Compulsive Disorder, 7% personality disorder, 4% simple phobia, 4% adjustment disorder, and 2% Primary Insomnia. Twenty-two clients had at least one comorbid diagnosis, while five had no formal diagnoses at all. Clients came from a variety of cultural backgrounds, education levels, and marital statuses. CMHR clients are referred elsewhere for services based on a limited set of exclusion criteria, including significant suicidality, legal involvement, current substance abuse, and active psychotic disorders. This sample of cases can be considered representative of the broader population of clients seen at this clinic.

Measures

Collaborative Case Conceptualization Rating Scale and Coding Manual (CCC-RS): Developed by Padesky, Dudley, and Kuyken (2011), the CCC-RS is a manual, available online, that guides coders in the rating of therapists' deployment of the four main principles of the collaborative case conceptualization method in either recorded or live therapy sessions. The

CCC-RS contains 14 items roughly split amongst the four sub-scales/main principles. Each item is accompanied by detailed coding instructions. Item scores range from a low of 0 to a high score of 3. Given that the CCC-RS originally coded therapy sessions, modifications were required to several items in order to better match with the content present in therapy reports. Specifically the CCC-RS manual gives examples of behaviours that can be observed within sessions (i.e. "the therapist expresses a high degree of curiosity, interest, and ***detailed*** questions..." pp. 17) which were adapted to reflect more overall trends that could be coded from reports ("the report evidences a high level of curiosity and dedication to understanding the client...").

Case Formulation Content Coding Method v.1 (CFCCM): Developed by Eells, Kendjelic, and Lucas (1998), this method for assessing case conceptualizations in reports was generated to be applicable across several psychotherapy orientations. It was found to have good reliability (Kendjelic & Lucas. 1998) and scores on the CFCCM have been shown to be positively related to therapist experience (Eells, Lombart, Kendjelic, Turner, & Lucas, 2005). The CFCCM v.1 has both content categories, relating to degree to which various possible types of information are present within a report, and several quality ratings. Quality ratings (from 0-5) are given to four main content categories (the common factors of case conceptualizations); symptoms and problems, precipitating stressors, predisposing life events, and inferred mechanism for linking the previous three categories and explaining client's current difficulties, which the developers found included across the majority of case conceptualization approaches.

Additionally, ratings are made on the overall quality of the conceptualization in the following areas: complexity of the conceptualization; how inferential the conceptualization is versus being merely descriptive; and how precise and tailored the language is in the conceptualization. For the purposes of this study, several changes were made to the CFCCM to

simplify the coding, eliminate content categories that were predicted to have an extremely low base rate, be more comparable to scoring on the CCC-RS, and to re-orient some of the coding to focus more on the quality and integration of the content categories instead of focusing solely their degree of presence. In our modified version, each remaining content category received a score from 0 to 3. A score of 0 indicated that particular category was absent, while a score of 1 indicated the content was present in a very limited or ambiguous form without integration with other elements of the conceptualization. A score of 2 indicated at least one clear mention of this content category with adequate integration with the rest of the conceptualization in a way that may inform the direction of therapy. A score of 3 was given for outstanding examples of a content category such that it was clear how this information being included could guide therapy with good links to other elements of the conceptualization. Finally, similar to other research with the CFCCM (Nattrass, Kellett, Hardy, & Ricketts, 2014) an overall quality item was generated based on a coder's overall impression of the conceptualization based on the previous elements of the CFCCM.

Additional Items coded: In addition to the two main coding methods described above, several other variables of interest were rated by our coders. Three therapeutic alliance items were generated to tap into the three alliance subscales found within the Working Alliance Inventory (Horvath & Greenberg, 1989) and Bordin's (1980) tripartite model of the alliance. See Table C4 within Appendix C for descriptions of these items. Scores ranged from a low of 1 to a high of 5 for each. Our team of coders demonstrated high inter-rater reliability in coding for these alliance items. Coders also made ratings of the client's functioning at both the end of the assessment phase (before treatment), and at the time of the discharge report (after treatment). Scores ranged from 1, very low functioning, to 5, very high functioning. These ratings were made based on a

combination of the described intensity of client symptoms at the time of the report, the degree of impairment and distress experienced by the client, and scores on measures of the client's quality of life, interpersonal functioning, and intensity of symptoms (see Table B1, Appendix B, for ICCs, item means, and standard deviations of the alliance and pre-and-post treatment functioning items).

Intervention

The therapists in this study were clinical psychology students ranging from their third year of clinical training through to therapists at the clinic for placement in their final internship year. All students received weekly supervision from registered clinical psychologists. The majority (90%) of cases were approached primarily from a cognitive behavioural (CBT) orientation; however, approximately one quarter of cases included secondary therapy modalities (such as interpersonal therapy, Dialectical Behavioural Therapy, mindfulness/self-compassion, or problem solving/supportive work). Four therapy cases were not primarily CBT in orientation, being primarily one of each of the following: assertiveness training, psychodynamic therapy, Interpersonal therapy, or psycho-educational. Within these cases, elements of CBT were still present, and so were left in the sample.

Therapy provided at the clinic is not manualized therapy, though techniques from treatment manuals are sometimes incorporated. Before active treatment begins, several sessions are spent assessing the client's difficulties after which an assessment report is written and a conceptualization is produced and shared with the client. Following the end of therapy, a discharge report is written which generally recaps some information from the assessment and then documents the therapeutic approach, the client's progress, and additional information or

insights gathered over the course of treatment. The duration of therapy ranged from 1 to 28 sessions (M = 12.80, SD = 7.64). Forty-two percent of cases represented unplanned endings, where therapy was terminated by the client prior to completion of their treatment. Cases were identified for inclusion in this study after their termination, and as such they can be considered naturalistic examples of therapy at this training clinic, no modifications to treatments were made for cases included in this study.

Procedure

1) Adapt and Prepare Coding Methods: As mentioned previously, modifications to the coding methods were made to ensure their applicability to case reports while maintaining focus on the constructs of interest in the original coding schemes.

2) Select and prepare case reports: The 46 files that met inclusion criteria were accessed and any assessment, progress, or discharge reports within were copied with client names and birthdates, and therapist and supervisor names redacted.

3) Train team of coders: The team of coders consisted of two second year master's level students in clinical psychology and three undergraduate students in psychology. Training first involved introducing the undergraduate students to the broader structure and tools of CBT therapy (assessment and treatment planning, automatic thoughts, core beliefs, thought records, behavioral experiments, exposure therapy, etc.) as well as to other important concepts being coded in our study (such as the working alliance, therapy engagement, therapy dropout, and case complexity). Second, coders met over several weeks to discuss the two coding methods and the additional coding items utilized in the study. Each item's description was reviewed together and items which remained unclear to any coder were identified and further information was added to the

manual to guide coding. As well, examples of how each item might be represented in case reports were discussed. The final step in training was to meet over several weeks to code practice reports as a team until coders felt confident in their ability to independently code the sample of files. Within this training files were coded independently and disagreements in item scores were resolved through discussion. In addition to group training sessions, undergraduate coders were encouraged to contact the first author individually for more information on psychotherapy concepts and procedures if the content or concepts within any report was unclear.

4) Code the reports and assess reliabilities: Following training, the 46 sets of reports were then coded independently over the course of six weeks. Both assessment and discharge reports for each case received independent scores for each item of the CCC-RS and the CFCCM. Alliance ratings were made on the therapy case as a whole. Meetings were held weekly to discuss the general progress of coding, identify and discuss whether any particular items were presenting challenges, and to help ensure coder drift did not occur. Halfway through coding the corpus of reports reliabilities for items were calculated and brought to the weekly meeting, items with low reliability were reviewed and methods to improve consistency of ratings were identified.

Data Analysis

Following the completion of coding intraclass correlation coefficients (ICCs) (Shrout & Lane, 2012) were calculated on the entire dataset for each numerical/scalar variable coded. Fleiss' Kappa statistics were calculated for dichotomous present/absent items of the CFCCM. SPSS v.22 software was utilized for all statistical analyses. Although progress reports had been coded, they were excluded from data analysis as they were present for only 5 (11%) cases.

Item reliabilities were calculated twice, first with all five coders, then subsequently retaining only the three raters who as a group achieved highest reliability. Within a large majority of items the dropping of two undergraduate coders improved reliabilities; graduate coders were dropped from only four items (7%) across both assessment and discharge report coding. Following the dropping of coders, average ICC reliability across the CCC-RS items at assessment increased from .53 to .60, and for the CFCCM from .54 to .61. These reliabilities fall in the fair to moderate range (Shrout & Lane, 2012).

Following these steps, item 4 of the CCC-RS remained below acceptable cut-offs for reliability and was dropped from the CCC-RS levels subscale. In addition, internal reliability analyses for the collaboration subscale of the CCC-RS indicated that item six could be dropped as it failed to significantly correlate with the other two items of the subscale, likely due to a floor effect. For the remaining items scores were then averaged across raters and then subscales were re-calculated excluding the missing items. Item 12 of the CFCCM remained problematically unreliable and was excluded from CFCCM related analyses.

Results

Table 4 contains the zero-order correlations of the case CCC-RS subscales, representing study IVs, with the alliance and client-functioning items, which represent study DVs. Additionally, correlations were calculated between the three items assessing the goal, task, and bond components of the alliance. These correlated very highly and in fact, after correcting for attenuation due to imperfect inter-rater reliability of the items, correlations reached the maximum (see Table 5). This suggested that a total alliance score would be most appropriate in subsequent analyses.

An initial hierarchical linear regression was calculated to assess whether the CCC-RS could predict post-treatment functioning. The initial regression model of assessment ratings from the CCC-RS subscales (Levels, Collaboration, Empiricism, and Strengths Focus) was not significant, F (4, 39) = 2.02, p = .11, R^2 = .17. Step two added the discharge scores for the same CCC-RS subscales. This model was significant, ΔF (4, 35) = 4.87, p = .003, ΔR^2 = .30, and accounted for 47% of the variation in post-treatment functioning. As the CCC-RS predictors were highly inter-correlated (see Table 4), the effect of adding the set of predictors to the regression equation should be focused on, while the interpretation of individual regression coefficients should be eschewed (as they will be unstable due to colinearity). Regardless, full details regarding this regression can be found within Appendix B, under Table B2.

A related and relevant analysis to the previous regression was to test the effect of reversing the order in which steps were entered. Within this regression the first step entered in the four CCC-RS subscales coded at discharge. The resulting model was significant, F (4, 39) = 5.45, p = .001, R^2 = .37. Next the CCC-RS subscales from assessment coding were entered, but their addition did not improve the model significantly, ΔF (4, 35) = 1.67, p = .18, ΔR^2 = .10. Details of this regression can be found in Appendix B, Table B3.

Pre-and-post treatment functioning were strongly related (r = .68, p<0.01). This suggests that the predictive ability of discharge scores for the CCC-RS subscales on post-treatment functioning might be due to pre-treatment functioning. Pre-treatment functioning may be a proxy variable for a case's complexity, given how this item was coded. As such, pre-treatment functioning may be an important control variable to account for the possible effect more complicated cases may have on the quality of a case conceptualization. Following this logic, a second hierarchical linear regression was performed with pre-treatment functioning entered as a

control variable. In step one, pre-treatment functioning was added which produced a significant model, F (1, 42) = 36.10, $p < .001$, $R^2 = .46$. The addition of the assessment CCC-RS subscale scores did not significantly improve the model in step two, ΔF (4, 38) = .60, $p = .66$, $\Delta R^2 = .03$. In the final step of the regression, discharge ratings of the four CCC-RS subscales were added, significantly improving the model, ΔF (4, 34) = 8.11, $p < .001$, $\Delta R^2 = .25$, which at this step accounted for 74% of the variance in post-treatment functioning. Details of this regression can be found in Appendix B, Table B4.

To examine the predictive power of the CCC-RS scores on the therapeutic alliance, a further hierarchical regression analysis was performed. As in the previous analyses, assessment ratings from the four CCC-RS were added in step one of the regression. This did not result in a significant model, F (4, 39) = .16, $p = .96$, $R^2 = .02$. Addition of the discharge report coding of the CCC-RS subscales did produce a significant regression model, ΔF (4, 35) = 17.64, $p < .001$, $\Delta R^2 = .66$, which accounted for 67% of the variation in overall alliance scores. Details of this regression can be found in Appendix B, Table B5.

To reiterate, one major goal of this study is to evaluate the relationship between the CCC-RS and treatment outcome. As previously described, the subscales of the CCC-RS appeared to inter-correlate to an extent that interpreting the coefficients within our regression analyses would be inappropriate. This complicated our ability to examine relationships between CCC-RS subscales, therapy outcome, and the alliance within our regression analyses. Given this, an examination of the full set of correlations between the IVs and DVs of the study was conducted to better elucidate the results of the regression analyses so far. As displayed in Table 4, scores from all four of the discharge report CCC-RS subscale scores significantly and positively correlated with post-treatment functioning, with correlations ranging from .39 to .58. These

correlations were re-examined as partial correlations controlling for pre-treatment functioning. Within these first order correlations the levels subscale (r = .55, p < .001), collaboration subscale (r = .60, p < .001), empiricism subscale (r = .50, p = .001), and strengths focus subscale (r = .55, p < .001) still showed strong positive relationships with post-treatment functioning. Two CCC-RS subscales from the assessment report coding, levels of conceptualization and collaboration, also significantly correlated with post-treatment functioning (zero-order correlations), but these correlations were negative in direction. Additionally, the empiricism subscale from assessment approached a significant negative correlation at r = -.27, p = .08. Scores for the therapeutic alliance positively correlated with both pre-and-post treatment functioning and with only the discharge report CCC-RS ratings (see Table 4).

Finally, pre-treatment functioning was also significantly and negatively correlated with the collaboration (r = -.44, p < .01) and empiricism (r = -.30, p = .04) subscales of the CCC-RS coded from the assessment reports. The levels subscale also approached a significant negative correlation at r = -.28, p = .06. In order to better ascertain the relationship between assessment CCC-RS scores and pre-treatment functioning, a hierarchical linear regression was performed. Assessment CCC-RS subscales entered into step one of this regression produced a significant model, F (4, 39) = 3.91, p = .01, R^2 = .29. The addition of the CCC-RS subscale scores from discharge report coding did not significantly improve the model, ΔF (4, 35) = .22, p = .92, ΔR^2 = .02. Details of this regression can be found in the Appendix B, Table B6.

Associations between CFCCM scores and our DVs were also explored (See Table 6 for zero-order correlations). As the CFCCM contains no theoretically prescribed subscales, two composite subscales were generated based on the groupings of items within the CFCCM itself. The first subscale consists of the "common factor" items of the CFCCM, as outlined previously.

The second subscale was calculated from the overall quality items of the CFCCM, excluding item 12 due to poor inter-rater reliability. Lastly we also calculated a total score to represent the breadth of information within conceptualizations. This subscale was calculated by summing each subcategory score from the CFCCM, where each subcategory score represented the average of our coders' present/absent scores. These three scales were then entered into a regression analysis similar to those previously conducted.

In step one the pre-treatment functioning of clients was entered as a control variable. Step two added the assessment CFCCM subscales just described. This model was not significantly better than the previous step, ΔF (3, 39) = .1.63, p = .20, ΔR^2 = .06. The addition of the three CFCCM scales from discharge coding did result in a significant improvement, ΔF (3, 36) = 5.91, p = .002, ΔR^2 = .16. The final model accounted for 62% of the variance in post-treatment functioning. Examinations of the bivariate correlations between these subscales, pre-and-post treatment outcome and the therapeutic alliance indicate some similarities to the CCC-RS subscale correlations; assessment overall quality ratings were significantly and negatively correlated with pre-treatment functioning (r = -.42, p = .004) and post-treatment functioning (r = -.47, p = .001). However, for the discharge CFCCM scales only the content total scale significantly correlated with post-treatment outcome, and did so negatively (r = -.37, p = .01). The CFCCM subscales generally did not significantly correlate with the alliance scores, with the exception being the scale representing the common factor item quality in the discharge coding, which did positively correlate with the alliance (r = .49, p = .001).

One observation that had been made by the team of coders was that for cases where an unplanned therapy termination had occurred discharge reports were often shorter and contained less detail. This suggested that if a relationship was being seen between discharge report

conceptualization scores and post-treatment functioning, it may simply be due to longer reports being written for therapy completers, who in turn have better treatment outcomes due to longer or more successful treatments. To address this possibility, therapy ending status (planned vs. unplanned endings) were dummy coded and entered into a final hierarchical regression analysis predicting post treatment functioning. To preserve statistical power with within our sample size a CCC-RS total score was included in the regression in lieu of the four CCC-RS subscales.

Pre-treatment functioning was entered again as a control variable in the initial step of this regression. In step two the planned versus unplanned ending dummy codes were entered, significantly improving the model, ΔF (1, 41) = 42.41, $p < .001$, $\Delta R^2 = .27$. In the subsequent step two variables, the CCC-RS total for the assessment and the CCC-RS total score at discharge, were added. The model again significantly improved, ΔF (2, 39) = 3.61, $p = .03$, $\Delta R^2 = .04$, with a total of 77% of variance in post-treatment functioning explained. This suggests that even after accounting for the variance in outcome attributed to planned or unplanned ending, conceptualization quality continued to predict post-treatment functioning. Discharge CCC-RS total scores appeared to drive this improvement, demonstrating a significant and positive relationship to post-treatment functioning ($\beta = .27$, $p = .01$) while assessment CCC-RS total scores did not show any significant relationship ($\beta = -.08$, $p = .33$). The final step of the regression included interactions between conceptualization quality and ending status. This did not significantly improve the model, failing to support a potential interaction between ending status and conceptualization quality predicting post-treatment functioning. Details of this regression can be found in the Appendix B, Table B7.

Table 4: Bivariate Correlations of CCC-RS subscales, Client Functioning, and the Alliance

	1	2	3	4	5	6	7	8	9	10	11
1) Pre-T_x Functioning	-										
2) Post-T_x Functioning	.68**	-									
Assessment Reports											
3) CCC-RS Levels	-.28	-.33*	-								
4) CCC-RS Collab.	-.44**	-.37*	.78**	-							
5) CCC-RS Empiric.	-.30*	-.27	.83**	.72**	-						
6) CCC-RS Strengths	.15	.05	.22	.28	.19	-					
Discharge Reports											
7) CCC-RS Levels	-.02	.39**	.15	.09	.16	-.04	-				
8) CCC-RS Collab.	.04	.47**	.03	.10	.04	-.02	.71**	-			
9) CCC-RS Empiric.	-.03	.34*	.17	.17	.21	-.08	.78**	.74**	-		
10) CCC-RS Strengths	.28	.58**	-.24	-.21	-.17	.41**	.39**	.58**	.35*	-	
11) Alliance Total	.31*	.79**	-.03	-.03	.01	.06	.60**	.76**	.66**	.63**	-

Note: *Correlation is significant at the .05 level. **Correlation is significant at the .01 level

Table 5: Bivariate Correlations of Alliance Subcomponents

	1	2	3
1) Alliance Goal Item	-	1†	1†
2) Alliance Task Item	.91**	-	1†
3) Alliance Bond Item	.86**	.87**	-

Note: *Correlation is significant at the .05 level. **Correlation is significant at the .01 level
Note: † = Disattenuated correlations

Table 6: Bivariate Correlations for CFCCM Derived Subscales, Client Functioning, and the Alliance

	1	2	3	4	5	6	7	8	9
1) Pre-T_x Functioning	-								
2) Post-T_x Functioning	.68**	-							
Assessment Reports									
3) CFCCM 1-4	-.20	-.21	-						
4) CFCCM Quality Items	-.42**	-.47**	.67**	-					
5)CFCCM Content Total	-.19	-.24	.36*	.25	-				
Discharge Reports									
6) CFCCM 1-4	-.06	.29	-.01	.00	.01	-			
7) CFCCM Quality Items	-.31*	-.11	.15	.36*	.15	.70**	-		
8)CFCCM Content Total	-.34*	-.37*	-.02	.21	.22	.30	.43**	-	
9) Alliance Total	.31*	.79**	-.06	-.17	-.22	.49**	.24	-.21	-

Note: *Correlation is significant at the .05 level. **Correlation is significant at the .01 level
Note: Item 12 of the CFCCM, precision of language, was excluded due to poor inter-rater reliability.

Discussion

The main goal of this study was to explore relationships between the quality of conceptualizations within therapy reports to therapy outcome and the therapeutic alliance. Two methods for evaluating case conceptualizations were utilized for this purpose, the CCC-RS and the CFCCM. Previous research had suggested that case conceptualizations may have a positive impact on the alliance (Nattrass et al., 2014), though no impact of case conceptualization quality on treatment outcome was found when utilizing the CFCCM. Other research has shown mixed results regarding any conceptualization-outcome link (Shulte, Künzel, Pepping, & Shulte-Bahrenberg, 1992; Chadwick, Williams, & Mackenzie, 2003; Persons, Roberts, Zalecki, & Brechwald, 2006). In this study, the CCC-RS was adapted to match the source of information at hand, which consisted of therapy assessment and discharge reports; as such, the present study represents a novel expansion of the research that can be conducted from the collaborative case conceptualization approach.

We tested the hypothesis that CCC-RS rated case conceptualization quality would positively relate to post treatment functioning. To do so, we conducted a hierarchical regression including first the assessment and then discharge ratings for the four CCC-RS subscales (Levels, Collaboration, Empiricism, and Strengths Focus). The total model accounted for a significant amount of post-treatment variance ($R^2 = .35$). However, results suggested that the assessment scores were poor predictors. Indeed, discharge CCC-RS subscale scores entered into a second regression model accounted for 37% of variance alone, and this model was not significantly improved by the addition of the assessment CCC-RS ratings. This pattern remained when controlling for controlling for pre-treatment functioning and when planned versus unplanned ending was taken into account. Against our hypotheses, several correlations between assessment

CCC-RS ratings actually associated negatively with pre-and-post treatment functioning. Relationships between the CCC-RS items and the therapeutic alliance supported our hypothesis that an alliance-conceptualization link would be found. This was evidenced by strong positive correlations found between each CCC-RS subscale and the alliance. However, counter to our expectations these correlations were only found at discharge.

Also counter to our expectations, CFCCM scores did not appear to be strong predictors of post-treatment functioning. Assessment CFCCM scales representing the quality of the common factor items (symptoms and problems, precipitating stressors, predisposing factors, and inferred mechanism), the overall quality of conceptualizations, and the breadth of content did not produce a significant regression model predicting post-treatment functioning when controlling for pre-treatment functioning. The addition of these same scales as coded from the discharge reports did significantly increase the amount of post-treatment variance explained, accounting for an additional 15%; however, examination of the zero-order correlations between these scales and client functioning suggest that, in fact, the only scale that significantly associated with outcome, breadth of content, predicted worse outcomes as the breadth of content increased.

Given that case conceptualizations are often described as important tools for guiding therapy towards the most efficient and beneficial approach (Persons, 2006), it was expected that assessment CCC-RS and CFCCM scores would correlate with and predict post-treatment outcome to a greater extent than the CCC-RS and CFCCM scores obtained from discharge reports. Instead, the correlations suggest that a deeper, more collaborative, more empirically rigorous conceptualization, with a higher overall quality, at assessment, is associated with a lower post-treatment level of client functioning.

Upon further reflection this appears reasonable: given a more challenging case a therapist may be forced to gather more information, conceptualize more thoroughly, and may make extra efforts to be collaborative and practice from an empirically solid framework. This seems to conflict with research showing that in cases with higher complexity, a lower conceptualization quality was found and that therapists' conceptualizations often did not guide therapy as much as overt client symptoms (Groenier, Pieters, Witteman, & Lehmann, 2014). However, the methodology in that study differed from ours in that therapist conceptualizations and treatment suggestions were generated in response to two patient vignettes. It is possible that in the context of real-world clinical work, more time or effort is brought to developing and articulating a complicated conceptualization when presented with a more challenging case and/or that the collaborative nature of in-vivo assessments may indeed improve case conceptualization quality for complex cases. Despite these efforts, case complexity may be negatively associated with client outcome. In our analyses, assessment CCC-RS subscale scores did fail to produce a significant model predicting post-treatment functioning, even when controlling for level of pre-treatment functioning. This suggests that even after accounting for the influence of case complexity and its relation to assessment CCC-RS scores, little relationship can be found between the conceptualization generated early in treatment and therapy outcome.

As mentioned previously, higher case conceptualization quality, as determined by the CCC-RS discharge scores, did significantly predict post-treatment functioning. This pattern was found both in both our raw correlations and across our regression models, even after controlling for pre-treatment functioning. CFCCM scores at discharge also appeared to predict post-functioning and this effect was likely driven by the negative relationship between increased breadth of information in the conceptualization and client functioning. This correlation may be

due to some of the subcategories of this scale specifically dealing with factors that could reasonable temper the likelihood of treatment success, such as negative treatment expectations, family history of psychiatric involvement, and multiple mechanisms explaining the client's difficulties. It may be that a therapist's awareness of possible risk factors for negative treatment outcomes does not always allow for them to overcome them.

One possible explanation for the relationship between higher CCC-RS quality at discharge and higher post-treatment functioning is that therapists whose clients made significant gains may write longer, more elaborated discharge reports due to feeling more positive about the case and having more engagement and interest in articulating this success. In contrast, clinicians may write less elaborate reports for clients who drop out of therapy or remain but make little progress. Writing a discharge report following less successful therapy may be a somewhat aversive experience and the resulting report may be more perfunctory in scope.

Additionally, it may be that within the context of a positive therapeutic alliance, which was found to significantly correlate with post-treatment outcome, more information is shared and a richer experience between client and therapist arises over therapy. This may translate to a more detailed and elaborated conceptualization and report, but as a somewhat spurious consequence of the therapeutic alliance, not directly attributable to the better conceptualization.

In an unpublished doctoral dissertation (Gower, 2011) which stemmed from early validation research into the CCC-RS it was found that the strengths focus subscale showed the strongest relationship to depressive symptom reduction. Within the present study, strengths focus also showed some unique characteristics amongst the CCC-RS subscales. First, although not a significant correlation, the strengths focus subscale at assessment was the only subscale at

assessment that did not have a negative relationship to pre-treatment functioning. Secondly, discharge strengths focus also had the strongest correlation with outcome, similar to the findings of Gower (2011). What cannot be ascertained however is the causal direction of these relationships. It may be that strengths present at in conceptualizations at assessment are merely the bi-product of less impairment and that at discharge a person who has generally improved more will also have more strengths to recognize or incorporate into treatment.

The final relationship of interest examined in this study was between the therapeutic alliance and collaborative case conceptualizations. Here the assessment report coding of CCC-RS and CFCCM subscales appeared unrelated to the overall alliance quality. This finding differs from the impact seen on the alliance by Nattrass and colleagues (2014), who reported increases in the working alliance following the formulation phase of therapy and concurrent with the sharing of a case conceptualization with a client. In the present study, the alliance rating was based on the impression gathered from both the assessment and discharge reports. This limits what conclusions can be made regarding the assessment conceptualization and alliance, particularly as the assessment reports represent work done with a client before a session is devoted to a thorough discussion of the formulation. It may be that sharing a quality case conceptualization following writing the assessment report would show an effect on the alliance but this resolution of temporal detail was unavailable.

The results of this study do suggest, however, that conceptualization quality is related to the alliance at least at the end phase of therapy. The causal direction of this relationship is also unclear, and given the inter-correlations between CCC-RS subscales it is difficult to determine precisely what elements relate to the alliance most strongly. Examination of the zero-order correlations do indicate that collaboration shares the strongest correlation to the alliance, which

may be due to the conceptual links between this subscale and elements of the alliance such as agreement on goals and having a positive, mutually respectful relationship. The CFCCM common factors at discharge also showed a positive relationship with the alliance, but why these items, and not the overall quality of the formulation or the breadth of information, were relevant to the alliance is unknown. It may be that focusing on the main drivers of a client's current problems, as opposed to creating complex or more inferential conceptualizations, is more beneficial to the alliance.

There are several limitations of this study. First, the study design was non-experimental, and as such drawing causal conclusions from the data is impossible. The significant correlations and regression models developed do not paint a clear picture of how the variables of interest are impacting each other. Second, these data represent only the therapist's impressions of the progress of therapy and the development of the case conceptualization. Although in many areas it appeared as though a reasonable inference could be made on the client's own contribution to the therapy or the case conceptualization process, these inferences are ultimately being made based on what the therapist has decided to include in these reports.

Related to this, it may be that in some cases the conceptualizations available within these reports were not complete. Case conceptualization may have occurred mainly off paper, either within the clinician's own mind, or in the supervisory dyad, and thus may not have been faithfully reflected in the notes available in the clinical file. It may also be that some hypotheses the therapist worked from or tested in therapy may not have been included in reports, particularly if they had little direct evidence or the hypotheses were somewhat sensitive in nature. In short, some therapists may have been hesitant to speculate too loosely in the context of a psychological report, particularly at the assessment phase.

As Kuyken, Padesky, and Dudley (2009) propose, case conceptualizations are likely to change and grow over time. As such, the conceptualizations within the assessment reports, produced before any active therapy techniques are deployed, may be fundamentally different from those generated after the therapy proper has begun and more active exploration, building, and testing of conceptualization hypotheses can be done. As such, the assessment conceptualizations may not be strongly reflective of the conceptualizations generated even shortly after the assessment reports are produced.

A therapist's overall writing style, the style of his/her supervisor, and the format of reports at the clinic also may have influenced this study. Within the CMHR, report format is guided by several sample reports made available to the trainee clinicians. These reports serve the functional role of recording assessment results, diagnoses, and disposition, but do not emphasize case conceptualization. To the extent that student therapists relied on the model reports it may have constrained the information provided in the report. One prominent area where this was the case was within discharge reports. Although some therapists used them to articulate a great deal of the therapy process and the conceptualization, others much more concisely noted a client's movement within therapy and their discharge status and referred readers to the assessment report for further information.

One final limitation is that novice therapists, such as the graduate level clinicians within this study, tend to produce treatment plans with poorer fits with their conceptualizations when compared to experts in case conceptualizations (Eells, Lombart, Kendjelic, Turner, & Lucas, 2005). This may temper the relationship between assessment conceptualization ratings and post-treatment outcomes. Related to this, due to the study's methodology we were not able to make direct judgments on how closely a therapist was guided by the case conceptualization in sessions.

Given these limitations and the results obtained within this study, it appears likely that a prospective study with repeated and direct examinations of conceptualizations, the therapeutic alliance, and symptoms changes, would be beneficial for several reasons. The first benefit may be that the temporal associations between CCC-RS quality and outcome or the alliance could be better explored. Second, through observations of therapy sessions, the possible confounds of coding from therapy reports may be minimized. Finally, additional outcome measures can be devised beyond coding an overall impression of client functioning. Alternatively, conducting experimental studies with random assignment between conceptualization driven conditions versus standardized conditions may be useful. Although research of this kind has been conducted in the past, the authors know of no such research available that have approached these designs from the perspective of collaborative case conceptualization.

Overall, it appears as though case conceptualization quality may indeed be a predictor of treatment outcome but this relationship is more nuanced than a universal association between higher quality conceptualizations and better treatment outcomes. Factors such as the complexity of the presenting problems and the stage of therapy at which the conceptualization is obtained may each contribute to this relationship and moderate the effect. This suggests the importance of controlling for the complexity and difficulty of cases, and ensuring a complete case conceptualization is obtained, when evaluating case conceptualizations in future research.

References

Adam, Y., Meinlschmidt, G., Gloster, A. T., & Lieb, R. (2012). Obsessive–compulsive disorder in the community: 12-month prevalence, comorbidity and impairment. *Social Psychiatry and Psychiatric Epidemiology*, *47*(3), 339-349.

Aston, R. (2009). A literature review exploring the efficacy of case formulations in clinical practice. What are the themes and pertinent issues?. *The Cognitive Behaviour Therapist*, *2*(02), 63-74.

Barber, J. P., & Crits-Christoph, P. (1993). Advances in measures of psychodynamic formulations. *Journal of Consulting and Clinical Psychology*, *61*(4), 574-585.

Beck, J. S. (1995). *Cognitive behavior therapy: Basics and beyond*. Guilford Press.

Benjamin, L. S. (2003). *Interpersonal reconstructive therapy: Promoting change in non-responders.* New York, NY: Guilford Press

Bieling, P. J., & Kuyken, W. (2003). Is cognitive case formulation science or science fiction?. *Clinical Psychology: Science and Practice*, *10*(1), 52-69.

Blackburn, I. M., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., & Reichelt, F. K. (2001). The revised cognitive therapy scale (CTS-R): psychometric properties. *Behavioural and Cognitive Psychotherapy*, *29*(04), 431-446.

Bordin, E. S. (1980, June). Of human bonds that bind or free. *Presented at the annual meeting of the Society for Psychotherapy Research, Pacific Grove, CA.*

Chadwick, P., Williams, C., & Mackenzie, J. (2003). Impact of case formulation in cognitive behaviour therapy for psychosis. *Behaviour Research and Therapy*, *41*(6), 671-680.

Chambless, D. L., Baker, M. J., Baucom, D. H., Beutler, L. E., Calhoun, K. S., Crits-Christoph, P., ... & Woody, S. R. (1998). Update on empirically validated therapies, II. *The Clinical Psychologist*, *51*(1), 3-16.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*(1), 7.

Crits-Christoph, P., Cooper, A., & Luborsky, L. (1988). The accuracy of therapists' interpretations and the outcome of dynamic psychotherapy. *Journal of Consulting and Clinical Psychology*, *56*(4), 490-495.

Dudley, R., Ingham, B., Sowerby, K., & Freeston, M. (2015). The utility of case formulation in treatment decision making; the effect of experience and expertise. *Journal of Behavior Therapy and Experimental Psychiatry, 48*, 66-74.

Dudley, R., Park, I., James, I., & Dodgson, G. (2010). Rate of agreement between clinicians on the content of a cognitive formulation of delusional beliefs: The effect of qualifications and experience. *Behavioural and Cognitive Psychotherapy*, *38*(02), 185-200.

Eells, T. D. (Ed.). (2011). *Handbook of psychotherapy case formulation*. Guilford Press.

Eells, T. D., Kendjelic, E. M., & Lucas, C. P. (1998). What's in a case formulation?: development and use of a content coding manual. *The Journal of Psychotherapy Practice and Research*, *7*(2), 144-153.

Eells, T. D., Lombart, K. G., Kendjelic, E. M., Turner, L. C., & Lucas, C. P. (2005). The quality of psychotherapy case formulations: a comparison of expert, experienced, and novice cognitive-behavioral and psychodynamic therapists. *Journal of Consulting and Clinical Psychology*, *73*(4), 579-589.

Eells, T. D., Lombart, K. G., Salsman, N., Kendjelic, E. M., Schneiderman, C. T., & Lucas, C. P. (2011). Expert reasoning in psychotherapy case formulation. *Psychotherapy Research, 21*(4), 385-399.

Emmelkamp, P. M., Bouman, T. K., & Blaauw, E. (1994). Individualized versus standardized therapy: A comparative evaluation with obsessive-compulsive patients. *Clinical Psychology & Psychotherapy, 1*(2), 95-100.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378-382.

Ghaderi, A. (2006). Does individualization matter? A randomized trial of standardized (focused) versus individualized (broad) cognitive behavior therapy for bulimia nervosa. *Behaviour Research and Therapy, 44*(2), 273-288.

Gower, P. (2011). *Therapist competence, case conceptualization and therapy outcome in cognitive behavioural therapy* (Doctoral dissertation, University of Exeter, Exeter, United Kingdom). Retrieved from https://ore.exeter.ac.uk/repository/handle/10036/3275

Groenier, M., Pieters, J. M., Witteman, C. L., & Lehmann, S. R. (2014). The Effect of Client Case Complexity on Clinical Decision Making. *European Journal of Psychological Assessment, 30(2),* 150-158.

Haarhoff, B. A., Flett, R. A., & Gibson, K. L. (2011). Evaluating the content and quality of cognitive-behavioural therapy case conceptualisations. *New Zealand Journal of Psychology, 40(*3), 104-114.

Haarhoff, B., Gibson, K., & Flett, R. (2011). Improving the quality of cognitive behaviour therapy case conceptualization: The role of self-practice/self-reflection. *Behavioural and Cognitive Psychotherapy*, *39*(3), 323-339.

Haynes, S. N., & Williams, A. E. (2003). Case formulation and design of behavioral treatment programs: Matching treatment mechanisms to causal variables for behavior problems. *European Journal of Psychological Assessment*, *19*(3), 164-174.

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*, *36*(2), 223-233.

Jacobson, N. S., Schmaling, K. B., Holtzworth-Munroe, A., Katt, J. L., Wood, L. F., & Follette, V. M. (1989). Research-structured vs clinically flexible versions of social learning-based marital therapy. *Behaviour Research and Therapy*, *27*(2), 173-180.

Kendjelic, E. M., & Eells, T. D. (2007). Generic psychotherapy case formulation training improves formulation quality. *Psychotherapy: Theory, Research, Practice, Training*, *44*(1), 66-77.

Kuyken, W., Beshai, S., Dudley, R., Abel, A., Görg, N., Gower, P., ... & Padesky, C. A. (2015). Assessing Competence in Collaborative Case Conceptualization: Development and Preliminary Psychometric Properties of the Collaborative Case Conceptualization Rating Scale (CCC-RS). *Behavioural and Cognitive Psychotherapy*, 1-14.

Kuyken, W., Fothergill, C. D., Musa, M., & Chadwick, P. (2005). The reliability and quality of cognitive case formulation. *Behaviour Research and Therapy*, *43*(9), 1187-1201.

Kuyken, W., Padesky, C. A., & Dudley, R. (2008). The science and practice of case conceptualization. *Behavioural and Cognitive Psychotherapy*, *36*(06), 757-768.

Kuyken, W., Padesky, C. A., & Dudley, R. (2009). *Collaborative case conceptualization: Working effectively with clients in cognitive-behavioral therapy*. Guilford Press

Mumma, G. H., & Smith, J. L. (2001). Cognitive–behavioral-interpersonal scenarios: interformulator reliability and convergent validity. *Journal of Psychopathology and Behavioral Assessment*, *23*(4), 203-221.

Nattrass, A., Kellett, S., Hardy, G. E., & Ricketts, T. (2014). The Content, Quality and Impact of Cognitive Behavioural Case Formulation During Treatment of Obsessive Compulsive Disorder. *Behavioural and cognitive psychotherapy*, *43*(5), 590-601.

Needleman, L. D. (1999). *Cognitive case conceptualization: A guidebook for practitioners*. Routledge.

Newman, M. G., Przeworski, A., Fisher, A. J., & Borkovec, T. D. (2010). Diagnostic comorbidity in adults with generalized anxiety disorder: Impact of comorbidity on psychotherapy outcome and impact of psychotherapy on comorbid diagnoses. *Behavior Therapy*, *41*(1), 59-72.

Padesky, C. A, Kuyken, W. & Dudley, R. (2011). *Collaborative Case Conceptualisation Rating Scale and Coding Manual*. Unpublished manual.

Persons, J. B. (1989). *Cognitive therapy in practice: A case formulation approach* (pp. 109-118). New York: Norton.

Persons, J. B. (1991). Psychotherapy outcome studies do not accurately represent current models of psychotherapy: A proposed remedy. *American Psychologist*, *46*(2), 99-106.

Persons, J. B. (2005). Empiricism, mechanism, and the practice of cognitive-behavior therapy. *Behavior Therapy*, *36*(2), 107-118.

Persons, J. B. (2006). Case formulation–driven psychotherapy. *Clinical Psychology: Science and Practice, 13*(2), 167-170.

Persons, J. B., & Bertagnolli, A. (1999). Inter-rater reliability of cognitive-behavioral case formulations of depression: A replication. *Cognitive Therapy and Research*, *23*(3), 271-283.

Persons, J. B., Mooney, K. A., & Padesky, C. A. (1995). Interrater relability of cognitive-behavioral case formulations. *Cognitive Therapy and Research*, *19*(1), 21-34.

Persons, J. B., Roberts, N. A., Zalecki, C. A., & Brechwald, W. A. (2006). Naturalistic outcome of case formulation-driven cognitive-behavior therapy for anxious depressed outpatients. *Behaviour Research and Therapy, 44*(7), 1041-1051.

Restifo, S. (2011). Beyond diagnosis-and therapy-centred therapy: the case for incorporating contextual factors in psychotherapy treatment planning. *Australasian Psychiatry*, *19*(4), 309-312.

Schulte, D., Künzel, R., Pepping, G., & Schulte-Bahrenberg, T. (1992). Tailor-made versus standardized therapy of phobic patients. *Advances in Behaviour Research and Therapy*, *14*(2), 67-92.

Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research, 7*(3), 301-317.

Shrout, P.E., & Lane, S.P. (2012). Reliability. In Cooper, H. E., Camic, P. M., Long, D. L., Panter, A. T., Rindskopf, D. E., & Sher, K. J. (Eds.), *APA handbook of research methods in psychology, (pp. 643 – 660). (Vol.1: Foundations, planning, measures,and psychometrics*). American Psychological Association.

Sim, K., Gwee, K. P., & Bateman, A. (2005). Case formulation in psychotherapy: Revitalizing its usefulness as a clinical tool. *Academic Psychiatry*, *29*(3), 289-292.

Appendix A: Supplementary Results From Study 1

Table A1. Inter-rater Reliabilities for CCC-RS and CFCCM items from Discharge Coding

CCC-RS	ICC (all coders)	ICC (lowest dropped)	Coders Dropped
Levels of Conceptualization Items			
1. Clear Link to Goals	.69	.70	Und. 2
2. Clear Rationale and Engagement	.62	.66	Und. 2
3. Meaningful Account of Issues	.67	.67	N/A
4. Good Parsimony*	.33	.47	Und. 1, Und. 2
Collaboration Items			
5. Collaboratively Developed CC	.81	.81	N/A
6. Culture and Experience*	.26	.67	Und. 1, Und. 2
7. Genuine Curiosity	.46	.55	Und. 2
Empiricism Items			
8. Justified CBT model	.72	.72	N/A
9. Test of "fit" of conceptualization	.43	.50	Und.1, Und.2
10. Treatment linked to CC	.70	.70	N/A
Strength and Resilience Focus Items			
11. Interest in Client Strengths	.71	.71	N/A
12. Client Strengths and Treatment	.59	.60	Und. 1
13. Client Aspiration Focus	.58	.58	N/A
14. Client Resilience Focus	.56	.56	N/A
Average ICC CCC-RS items	.58	.64	

CFCCM	ICC (all coders)	ICC (lowest dropped)	Coders Dropped
Content Quality Ratings			
1. Symptoms and Problems	.56	.68	Und. 2
2. Precipitating Stressors & Events	.57	.62	Und. 3
3. Predisposing Life Events	.70	.75	Und. 1
4. Inferred Mechanism	.63	.70	Und. 1, Und. 2
5. Client History Categories	.62	.65	Und. 3
6. Iatrogenic Factors	.61	.63	Grad. 2
7. Global Level of Adjustment	.67	.67	N/A
8. Treatment Indicators	.40	.40	N/A
9. Therapist T_x Expectations	.64	.71	Und. 2
Overall Quality Ratings			
10. Complexity of Formulation	.56	.60	Und. 1, Und. 2
11. Degree of Inference	.50	.60	Und. 1, Und. 2
12. Precision of Language	.53	.70	Und. 1, Und 2
13. Overall Formulation Quality	.68	.69	Und. 2
Average ICC CFCCM items	.59	.65	

Table A2. Inter-rater Reliabilities for CFCCM Present/Absent items (Discharge Coding)

	Fleiss' Kappa (all coders)	Fleiss' Kappa (lowest dropped)	Coders Dropped
Inferred Mechanism			
a. Inferred Mech. Psychological	-.07*	.15*	Und. 2
b. Inferred Mech. Biological	.32*	.32*	N/A
c. Inferred Mech. Socio-Cultural	.24*	.24*	N/A
d. Inferred Mech. Substance Use	.63	.63	N/A
Client History Categories			
a. Own or Family Psych. history	-.03	-.03	N/A
b. Own or Family Medical History	.04*	.04*	N/A
c. Developmental or Social History	.15*	.15*	N/A
Treatment Indicators			
a. Negative Treatment Motivation	.40	.54	Und. 1, Und. 2
b. Positive Motivation for Treatment	.22	.22	N/A
c. Positive Social Support	.28*	.35*	Und. 1
d. Posit. self perception, goal, wish	.07	.11	Und. 3, Grad. 1
Therapist Treatment Expectations			
a. Negative Treatment Indications	.52	.52	N/A
b. Prognosis	.01	.01	N/A
Average Fleiss' Kappa	.21	.25	

Table A3. Internal Reliabilities for Coding Methods and Subscales (Discharge Coding)

	Cronbach's α	
	All Coders	Lowest Dropped
Levels of Conceptualization Subscale	.89	.91
Levels of Conceptualization (Item 4 dropped)	.86	.88
Collaboration Subscale	.54	.56
Collaboration Subscale (Item 6 dropped)	.71	.78
Empiricism Subscale	.81	.77
Strength and Resilience Subscale	.90	.89
CCC-RS Total Scale (all items)	.92	.91
CCC-RS Total (Items 4 and 6 dropped)	.92	.91
	Composite Scale Reliability	
CFCCM Items 1-4	.79	.84
CFCCM Overall Items	.80	.83

Appendix B: Supplementary Results From Study 2

Table B1. Inter-rater Reliabilities and Descriptive Statistics for Alliance Items and Rating of Pre-and-Post Treatment Functioning

	ICC	Mean	*SD*
Alliance Goal Item	.87	3.29	.83
Alliance Task Item	.90	3.21	.94
Alliance Bond Item	.83	3.54	.71
Pre-functioning	.81	2.81	.61
Post-Functioning	.89	3.61	.79

Table B2. Hierarchical Linear Regression Predicting Post-T_x Functioning

	t	*p*	*β*	*ΔF*	*df*	*p*	R^2
Assess. Coding (Step 1)				2.02	4,39	.11	.17
Assess. CCC-RS Levels	-.66	.51	-.19				
Assess. CCC-RS Collab.	-1.48	.15	-.35				
Assess. CCC-RS Empiric.	.41	.68	.11				
Assess. CCC-RS Stren.	1.13	.27	.17				
Discharge Coding (Step 2)				4.87	4,35	.003	.35
Assess. CCC-RS Levels	-.42	.68	-.11				
Assess. CCC-RS Collab.	-1.41	.17	-.31				
Assess. CCC-RS Empiric.	.17	.86	.04				
Assess. CCC-RS Stren.	.30	.77	.05				
Dis. CCC-RS Levels	.60	.55	.13				
Dis. CCC-RS Collab.	.77	.45	.19				
Dis. CCC-RS Empiric.	.32	.75	.07				
Dis. CCC-RS Stren.	1.25	.22	.28				

Note: Due to multi-colinearity between CCC-RS subscales Beta coefficients should be interpreted with caution

Table B3. Reversed Hierarchical Linear Regression Predicting Post-T_x Functioning

	t	*p*	*β*	*ΔF*	*df*	*p*	R^2
Discharge Coding (Step 1)				5.65	4,39	.001	.37
Dis. CCC-RS Levels	.68	.50	.15				
Dis. CCC-RS Collab.	.41	.69	.09				
Dis. CCC-RS Empiric.	-.01	.99	.00				
Dis. CCC-RS Stren.	2.94	.006	.47				
Assess. Coding (Step 2)				1.67	4,35	.18	.47
Dis. CCC-RS Levels	.60	.55	.13				
Dis. CCC-RS Collab.	.77	.45	.19				
Dis. CCC-RS Empiric.	.32	.75	.07				
Dis. CCC-RS Stren.	1.25	.22	.28				
Assess. CCC-RS Levels	-.42	.68	-.11				
Assess. CCC-RS Collab.	-1.41	.17	-.31				
Assess. CCC-RS Empiric.	.17	.86	.04				
Assess. CCC-RS Stren.	.30	.77	.05				

Note: Due to multi-colinearity between CCC-RS subscales Beta coefficients should be interpreted with caution

Table B4. Hierarchical Linear Regression Predicting Post-T_x Functioning controlling for Pre-T_x Functioning

	t	*p*	*β*	*ΔF*	*df*	*p*	R^2
Controls (Step 1)				*36.10*	*1,42*	*<.001*	.46
Pre- T_x Functioning	*6.01*	*<.001*	*.68*				
Assess. Coding (Step 2)				.60	4,38	.66	.49
Pre- T_x Functioning	4.92	<.001	.67				
Assess. CCC-RS Levels	-1.36	.18	-.31				
Assess. CCC-RS Collab.	.33	.75	.07				
Assess. CCC-RS Empiric.	.69	.49	.15				
Assess. CCC-RS Stren.	-.25	.81	-.03				
Discharge Coding (Step 3)				8.11	4,34	<.001	.74
Pre- T_x Functioning	5.99	<.001	.63				
Assess. CCC-RS Levels	-1.30	.20	-.24				
Assess. CCC-RS Collab.	.65	.52	.11				
Assess. CCC-RS Empiric.	.42	.67	.07				
Assess. CCC-RS Stren.	-1.15	.26	-.15				
Dis. CCC-RS Levels	1.45	.16	.22				
Dis. CCC-RS Collab.	.54	.60	.09				
Dis. CCC-RS Empiric.	.08	.94	.01				
Dis. CCC-RS Stren.	1.88	.07	.30				

Note: Due to multi-colinearity between CCC-RS subscales Beta coefficients should be interpreted with caution

Table B5. Hierarchical Linear Regression Predicting Alliance Total Score

	t	*p*	*β*	*ΔF*	*df*	*p*	R^2
Assess. Coding (Step 1)				.16	4,39	.96	.02
Assess. CCC-RS Levels	-.43	.67	-.14				
Assess. CCC-RS Collab.	-.37	.72	-.10				
Assess. CCC-RS Empiric.	.49	.63	.14				
Assess. CCC-RS Stren.	.35	.73	.06				
Discharge Coding (Step 2)				17.64	4,35	<.001	.67
Assess. CCC-RS Levels	-.03	.98	-.01				
Assess. CCC-RS Collab.	-.67	.51	-.12				
Assess. CCC-RS Empiric.	.29	.78	.05				
Assess. CCC-RS Stren.	-.34	.74	-.05				
Dis. CCC-RS Levels	.04	.97	.01				
Dis. CCC-RS Collab.	2.05	.05	.40				
Dis. CCC-RS Empiric.	1.44	.16	.25				
Dis. CCC-RS Stren.	1.75	.09	.31				

Note: Due to multi-colinearity between CCC-RS subscales Beta coefficients should be interpreted with caution

Table B6. Hierarchical Linear Regression Predicting Pre-T_x Functioning

	t	*p*	*β*	*ΔF*	*df*	*p*	R^2
Assess. Coding (Step 1)				3.91	4,39	.01	.29
Assess. CCC-RS Levels	.68	.50	1.83				
Assess. CCC-RS Collab.	-2.82	.01	-.62				
Assess. CCC-RS Empiric.	-.22	.83	-.06				
Assess. CCC-RS Stren.	2.14	.04	.30				
Discharge Coding (Step 2)				.23	4,35	.92	.30
Assess. CCC-RS Levels	.71	.48	.21				
Assess. CCC-RS Collab.	-2.67	.01	-.68				
Assess. CCC-RS Empiric.	-.18	.86	-.05				
Assess. CCC-RS Stren.	1.59	.12	.33				
Dis. CCC-RS Levels	-.59	.56	-.14				
Dis. CCC-RS Collab.	.55	.59	.15				
Dis. CCC-RS Empiric.	.37	.71	.10				
Dis. CCC-RS Stren.	-.14	.91	-.03				

Note: Due to multi-colinearity between CCC-RS subscales Beta coefficients should be interpreted with caution

Table B7. Hierarchical Linear Regression Predicting Post-T_x Functioning with Planned or Unplanned Ending Status and Overall Quality Scores

	t	*p*	*β*	*ΔF*	*df*	*p*	R^2
Controls (Step 1)				36.10	1,42	< .001	.46
Pre- T_x Functioning	6.01	>.001	.68				
Dummy Coded Ending Status (Step 2)				42.41	1,41	< .001	.74
Pre-T_x Functioning	7.82	>.001	.63				
Ending Status	6.51	>.001	.53				
CCC-RS Total (Step 3)				3.61	2,39	.04	.77
Pre-T_x Functioning	7.79	>.001	.61				
Ending Status	3.33	.002	.34				
CCC-RS Total Assess.	-.99	.33	-.08				
CCC-RS Total Discharge	2.67	.01	.27				
CCC-RS Total by Ending Status Interaction (Step 4)				.49	2,37	.75	.75
Pre-T_x Functioning	7.59	>.001	.60				
Ending Status	-.30	.77	-.15				
CCC-RS Total Assess.	-1.10	.28	-.11				
CCC-RS Total Discharge	.70	.49	.14				
Assess. CCC-RS by Ending Status	.56	.58	.24				
Discharge CCC-RS by Ending Status	.65	.52	.35				

Appendix C: Tables Relevant to Studies 1 and 2

Table C1: Overview of CCC-RS and CFCCM item constructs

CCC-RS	Subscale	Brief Description
1) Conceptualization linked to goals	Levels of Conceptualization.	The conceptualization is tied to the presenting issues, treatment goals, and priorities for therapy.
2) Clear rationale	Levels of Conceptualization.	There is a clear rationale for the case conceptualization and the elements within it, ideally it should be clear that the client engaged with and understood this conceptualization
3) Meaningful account	Levels of Conceptualization.	A meaningful account of the presenting issues has been made beyond simple lists or descriptions, and is matched to the clients ability to understand, and the stage of therapy
4) Good parsimony	Levels of Conceptualization.	The conceptualization is not so complex to hinder understanding or become convoluted, though it remains informative enough to help explain the key features of a client's case
5) Client and therapist collaboration	Collaboration	The therapist and client worked together to develop the conceptualization, both were engaged and added to the conceptualization in an open and interactive process.
6) Culture and Experience	Collaboration	Relevant aspects of a client's cultural experience were found in the conceptualization which also appeared to be tailored to be understood given the client's culture and background
7) Genuine curiousity	Collaboration	A high dedication to understanding the client can be found, unexpected information was welcomed into the conceptualization and a lack of presumptions and biases allowed for real insights to be made
8) Justified CBT model	Empiricism	The conceptualization appears to draw on psychological research and established models for psychopathology while remaining tied to the client's own experiences

9) Test of hypothesis "fit"	Empiricism	There is evidence that the conceptualization was tested for accuracy, either through detailed assessment, exploring alternative explanations, or through therapy exercises such as behavioural experiments
10) Treatment plan - conceptualization link	Empiricism	It is clear that the treatment plan was guided by the conceptualization and that the main foci of therapy were chosen based on the key features identified within the conceptualization
11) Interest in client strengths	Strengths and Resilience	Both client interests and strengths were included in the assessment and conceptualization, and the therapist may have helped the client discover hidden strengths
12) Strengths applied to treatment	Strengths and Resilience	The conceptualization and treatment plan utilized client strengths to promote a meaningful improvement in treatment outcome and resilience
13) Client aspirations	Strengths and Resilience	Interest in how the client would like things to be, above and beyond mere reductions in symptoms or distress, have been noted and included in the conceptualization and treatment plan
14) Client resilience	Strengths and Resilience	Client resilience is highlighted in order to help move the client towards treatment goals or increase the client's appreciation for their previous resilience and self-efficacy

CFCCM	**Sub-Category Items**	**Brief Description**
1) Symptoms and Problems		The signs, symptoms, and other clinically important phenomena the client is experiencing are described. Tying these to a diagnosis, noting their course or intensity, and the impact on client functioning indicate higher quality for this category
2) Precipitating Stressors		Events that have exacerbated or initiated the client's symptoms or problems are articulated.

3) Predisposing Life Events		Identification of specific traumatic life events or stressors that are assumed to increase a client's vulnerability to later development of their current difficulties
4) Inferred Mechanism	Psychological	Any psychologically based mechanism that explains the client's current difficulties
	Biological	A biological or genetic mechanism that has lead to the client's current difficulties
	Socio-cultural	Where ethnicity/acculturation, social economic status, or absence of social support has influenced the client's difficulties
	Substance abuse	Indications that a substance abuse or dependency has contributed to the client's difficulties
5) Client History Content Categories	Own or family psychiatric history	Information regarding the client's or their family's past experiences of psychiatric diagnoses or involvement
	Own or family medical history	Information regarding relevant medical issues within the client or their family's past
	Developmental and social history	Relevant information on the client's own developmental milestones or delays, and/or their early social experiences
6) Iatrogenic Factors		Relevant examples of negative impacts on the client as a result of previous mental health interventions
7) Global Level of Adjustment		Detailed and useful recognition of strengths or weaknesses within a clients daily functioning and level of overall impairment
8) Treatment Indicators	Negative treatment motivation	Signs of hesitance, reluctance, or mistrust, or lack of engagement in the treatment
	Positive motivation for treatment	Signs for high engagement, open and helpful contributions to therapy, and willingness to perform most therapy tasks

	Positive Social Support	Useful and beneficial social support from a client's family, friends, etc. that aide in the progress of therapy
	Positive aspects of client/goals/ self-perceptions	Aspects of the client that are healthy and well regulated, positive goals the client holds, or positive self-perceptions of the client are mentioned and incorporated into the formulation
9) Therapist Treatment Expectations	Neg. T_x indicators	Elements of the client's personality, history, attitudes, or beliefs which may negatively impact the progress of therapy are explored and included in the conceptualization to help avoid these potential barriers to progress
	Prognosis	The therapist makes predictions for the outcome of therapy as a result of information gathered and the match between the client's difficulties and the therapeutic approach
10) Complexity of conceptualization		The degree to which several facets of the client's current problems are explored integrated into a meaningful account within the conceptualization
11) Degree of inference		This is the degree to which the report goes beyond merely summarizing or describing the client's presenting problems and situation. Higher levels of inference include more internal psychological processes and hypothetical considerations
12) Precision of language		The degree to which the conceptualization and language within the report appear to describe a unique individual versus a generic language and descriptions
13) Overall Formulation quality		This is a rating for the overall quality of the formulation based on a consideration of all the previous items of the CFCCM together. It includes considering how well the conceptualization could guide therapy, integrate information, and help the therapist understand the client

Table. C2: Means and Standard Deviations for Items Across All 46 Assessment Reports

	M (*SD*) all coders	*M* (*SD*) lowest dropped
CCC-RS Items		
Levels of Conceptualization Items		
1. Clear Link to Goals	2.00 (.42)	2.14 (.54)
2. Clear Rationale and Engagement	1.98 (.37)	2.09 (.41)
3. Meaningful Account of Issues	1.96 (.40)	2.01 (.46)
4. Good Parsimony	1.76 (.29)	1.76 (.40)
Collaboration Items		
5. Collaboratively Developed CC	1.97 (.37)	1.98 (.46)
6. Culture and Experience	.58 (.53)	.57 (.59)
7. Genuine Curiosity	1.60 (.46)	1.60 (.51)
Empiricism Items		
8. Justified CBT model	1.69 (.52)	1.69 (.52)
9. Test of "fit" of conceptualization	1.39 (.45)	1.17 (.52)
10. Treatment linked to CC	1.97 (.44)	1.97 (.44)
Strength and Resilience Focus Items		
11. Interest in Client Strengths	.70 (.55)	.70 (.55)
12. Client Strengths and Treatment	.39 (.39)	.20 (.45)
13. Client Aspiration Focus	1.19 (.42)	1.19 (.42)
14. Client Resilience Focus	.51 (.38)	.42 (.49)
CCC-RS total score	-	19.47 (4.02)
CFCCM Items		
Content Quality Ratings		
1. Symptoms and Problems	2.37 (.30)	2.34 (.39)
2. Precipitating Stressors & Events	1.57 (.63)	1.52 (.72)
3. Predisposing Life Events	1.05 (.46)	.93 (.51)
4. Inferred Mechanism	1.77 (.43)	1.76 (.50)
5. Client History Categories	1.85 (.45)	1.85 (.45)
6. Iatrogenic Factors	.20 (.25)	.22 (.30)
7. Global Level of Adjustment	1.64 (.42)	1.64 (.42)
8. Treatment Indicators	1.44 (.46)	1.35 (.52)
9. Therapist T_x Expectations	1.05 (.53)	1.04 (.63)
Overall Quality Ratings		
10. Complexity of Formulation	2.96 (.59)	2.88 (.70)
11. Degree of Inference	2.73 (.64)	2.57 (.83)
12. Precision of Language	3.10 (.39)	3.09 (.57)
13. Overall Formulation Quality	3.17 (.49)	3.09 (.71)

Table. C3: Means and Standard Deviations for Items Across All 44 Discharge Reports

	M (*SD*) all coders	*M* (*SD*) lowest dropped
CCC-RS Items		
Levels of Conceptualization Items		
1. Clear Link to Goals	1.90 (.53)	1.93 (.53)
2. Clear Rationale and Engagement	1.84 (.50)	1.88 (.50)
3. Meaningful Account of Issues	1.68 (.48)	1.68 (.48)
4. Good Parsimony	1.72 (.47)	1.70 (.47)
Collaboration Items		
5. Collaboratively Developed CC	1.82 (.57)	1.82 (.57)
6. Culture and Experience	.16 (.36)	.14 (.36)
7. Genuine Curiosity	1.31 (.48)	1.40 (.48)
Empiricism Items		
8. Justified CBT model	1.68 (.63)	1.40 (.63)
9. Test of "fit" of conceptualization	1.25 (.60)	1.68 (.60)
10. Treatment linked to CC	1.88 (.45)	1.88 (.45)
Strength and Resilience Focus Items		
11. Interest in Client Strengths	.42 (.46)	.42 (.46)
12. Client Strengths and Treatment	.53 (.41)	.47 (.41)
13. Client Aspiration Focus	1.10 (.42)	1.10 (.42)
14. Client Resilience Focus	.42 (.38)	.42 (.38)
CCC-RS total score	-	17.99 (4.67)
CFCCM Items		
Content Quality Ratings		
1. Symptoms and Problems	1.62 (.45)	1.77 (.52)
2. Precipitating Stressors & Events	.58 (.48)	.59 (.56)
3. Predisposing Life Events	.25 (.37)	.13 (.33)
4. Inferred Mechanism	1.47 (.44)	1.75 (.57)
5. Client History Categories	.45 (.36)	.31 (.35)
6. Iatrogenic Factors	.47 (.45)	.44 (.45)
7. Global Level of Adjustment	1.38 (.54)	1.38 (.54)
8. Treatment Indicators	1.67 (.41)	1.67 (.41)
9. Therapist T_x Expectations	1.48 (.49)	1.38 (.61)
Overall Quality Ratings		
10. Complexity of Formulation	2.86 (.57)	2.72 (.77)
11. Degree of Inference	2.65 (.50)	2.48 (.66)
12. Precision of Language	2.90 (.50)	2.75 (.72)
13. Overall Formulation Quality	3.07 (.63)	2.98 (.73)

Table C4: Alliance Item Descriptions and Coding Guidelines for Coders

Alliance Goal Item: Did the client and therapist agree on goals for therapy?	Is there evidence of the client and therapist really agreeing on the goals for therapy? Did the therapist indicate they thought the client's goals were realistic and positive?
Alliance Task Item: Did the client feel the therapy was effective and well planned?	Is there evidence that the client thought the therapy was well planned and effective? Did the therapist note any comment from the client on how well things were going, or how much they enjoyed or appreciated their therapy? Did the therapist note any of the client's perceptions on the progress they were making?
Alliance Bond Item: Did the therapist and client like and respect each other?	Is there evidence that the relationship between the client and therapist was positive? Did the therapist note the friendliness and openness of the client towards them? Did the therapist note that the client and they had a good rapport? Was there an open and trusting relationship?

Table. C5: Percent Chance That a Coder Would Identify a CFCCM Subcategory as Present Within a Typical Report

CFCCM Subcategory	
Assessment Reports	
Inferred Psychological Mechanism	95%
Inferred Biological Mechanism	20%
Inferred Socio-cultural Mechanism	46%
Inferred Substance Abuse Mechanism	13%
Own or family psychiatric history	83%
Own or family medical history	83%
Developmental and social history	69%
Negative treatment motivation	13%
Positive motivation for treatment	58%
Positive Social Support	51%
Positive: aspects of client/goals/self-perceptions	49%
Neg. T_x indicators	25%
Prognosis	70%
Discharge Reports	
Inferred Psychological Mechanism	97%
Inferred Biological Mechanism	7%
Inferred Socio-cultural Mechanism	16%
Inferred Substance Abuse Mechanism	9%
Own or family psychiatric history	30%
Own or family medical history	3%
Developmental and social history	8%
Negative treatment motivation	58%
Positive motivation for treatment	74%
Positive Social Support	18%
Positive: aspects of client/goals/self-perceptions	33%
Neg. T_x indicators	50%
Prognosis	69%