CLIENTS' PERSPECTIVES OF THE WORKING ALLIANCE INVENTORY

JAMES WATSON-GAZE

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PSYCHOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

MAY 2017

Abstract

The working alliance is one of the most widely studied constructs in the psychotherapy research literature, and ratings of the working alliance are widely used in clinical research and practice to measure the relationships between clients and their therapists and courses of treatments. Nonetheless, relatively little is known about how clients understand and engage with the scales from which such ratings are derived. To address this gap in the literature, clients in ongoing psychotherapy were interviewed after completing a version of the most widely used alliance rating scale, the Working Alliance Inventory (WAI). This study was comprised of two primary phases. In the first phase, 12 clients took part in semi-structured interviews focused on understanding what they were aware of when completing a version of the WAI, the WAI – Short Revised (WAI-SR). Then, transcripts from this phase were analyzed using Grounded Theory methods. Analyses yielded a core category of Rolling on the Right Track, which encapsulated participant reflections of the scale as a means to communicate the degree to which clients and therapists are moving together towards shared therapeutic goals. In the second phase, 10 clients took part in cognitive interviews that targeted specific information about whether clients identified any difficulties with understanding elements of the WAI-SR. Transcripts were analyzed using the Classification Coding Scheme and findings indicated that participants noted difficulties with several parts of the scale. Item comprehension challenges were the most frequently raised areas of concern. Results were triangulated across both primary phases and key findings discussed in the context of ongoing debates in the research literature. This study exemplifies and underscores the dynamic nature of clients' rating processes when completing a version of the WAI, as well as the ways they

use the measure as a means to assess and communicate the degree to which they are working effectively in treatment. This study also raises concerns about the intelligibility of the WAI-SR for clients who complete it and calls into question the confidence with which researchers and clinicians can maintain their faith in the accuracy of clients' ratings.

Acknowledgments

I am incredibly grateful to the many people who helped this dissertation become a reality. This project would not have come to fruition were it not for the input from many people in my life.

First and foremost, thank you to the interview participants who volunteered their time to meet with me. It was a privilege to sit and learn from each of you.

To my supervisor, Lynne Angus, thank you for your guidance throughout my graduate school journey and this evolving research project. Thank you to my committee members - Henny Westra and John Eastwood - as well as the members of my examining committee – Karen Fergus, Livy Visano, and Tim Anderson. I greatly appreciate the time and care you took in reading and reflecting on this dissertation. A special thank you also to Faye Doell who helped to motivate me in this last leg of the race.

To my parents, you instilled in me a deep love of learning for which I am thankful every day. Your care and dedication over my many years of schooling have been indispensable. I could not have made it to this stage without your support and encouragement.

And to my love, Amanda – none of this would have meaning if I could not share it with you. Thank you for your support throughout this long process. I am grateful for each and every day we have together.

Table of Contents

List of Tables

List of Figures

Literature Review

Introduction

The working or therapeutic alliance (hereafter referred to as 'the alliance') represents the collaborative engagement between clients and therapists in psychotherapeutic treatments (Hatcher, 2010) and is one of the most widely cited common factors in psychotherapy research (Muran & Barber, 2010). It is a robust predictor of treatment outcome and may account for more outcome variance than any other psychotherapy process variable (Mallinckrodt & Tekie, 2016). A meta-analysis undertaken by Horvath, Del Re, Flückiger and Symonds (2011) found an average effect size of .28 when measuring the relationship between clients' ratings of their alliances and their therapeutic outcomes. While this relationship between alliance and outcome has been demonstrated consistently in the literature (Horvath & Bedi, 2002; Martin, Garske, & Davis, 2000), relatively little is known about how clients themselves understand their alliances, their alliance ratings, and the measures through which these ratings are acquired. The Working Alliance Inventory (WAI; Horvath & Greenberg, 1989) is the most widely used of the alliance rating scales, having been used in approximately 40% of the 201 studies identified by a recent meta-analysis (Horvath et al., 2011). Given the preponderance of the WAI in the literature and the paucity of understanding of clients' rating processes on this measure, the present study utilized two qualitative analytic methods in order to examine clients' approaches to and reflections on the items and rating processes utilized by a recent version of the WAI, the WAI - Short Revised (Hatcher & Gillaspy, 2006).

The following review presents findings from the clinical and research literatures on the working alliance and self-report measures. The first section reviews the literature on

the working alliance and describes the history of the alliance construct, as well as the history of the measurement of the alliance, and then finishes by outlining research findings on variables associated with or impacted by the alliance. The next section focuses on respondents' approaches to self-report scales in general, introduces the reader to survey response theories, and reviews research completed on survey methods. This review finishes with a description and rationale of the present study and provides an introduction to the research questions guiding the work.

Working Alliance

Horvath and Bedi (2002) offer the following as a comprehensive description of the alliance:

> The alliance refers to the quality and strength of the collaborative relationship between client and therapist in therapy. This concept is inclusive of: The positive affective bonds between client and therapist, such as mutual trust, liking, respect, and caring. Alliance also encompasses the more cognitive aspects of the therapy relationship: consensus about, and active commitment to, the goals of therapy and to the means by which these goals can be reached. Alliance involves a sense of partnership in therapy between therapist and client, in which each participant is actively committed to their specific and appropriate responsibilities in therapy, and believes the other is likewise enthusiastically engaged in the process. The alliance is a conscious and purposeful aspect of the relation between therapist and client. (cited in Horvath, 2001, p. 365)

In this definition, Horvath and Bedi (2002) highlight both the affective and cognitive elements of the alliance, together directed in the service of providing positive therapeutic change. Hatcher (2010) offers a related description:

> Across different therapeutic approaches, a good alliance means that patients and therapists are working well together towards the goals of therapy. Good work is expected to be purposeful and collaborative. Thus, the alliance is a way to think about how the patient and therapist are working together. (p. 8)

Hatcher further underscores the directed, goal-focused facets of the alliance, and later goes on to describe the alliance as a measure of the degree to which client and therapist are working together effectively within a given therapeutic orientation.

Horvath and Bedi's (2002) presentation of the alliance follows from Bordin's (1979) tripartite conceptualization of it as comprised of the affective bond established between clients and therapists, as well as the agreement between clients and therapists on the goals and the tasks of their courses of treatment. It is Bordin's model on which the Working Alliance Inventory (Horvath & Greenberg, 1989) is based and is thus the primary theoretical foundation for the present work.

A history of the alliance. The construct of the alliance can be traced to early writings on the therapeutic relationship in general, and Freud's writings on transference reactions in particular (Horvath, 2001; Safran & Muran, 2000). Freud (1912) understood the transference in psychoanalysis as a form of a patient's resistance to the analytic method and believed these reactions to be indicative of patients' unsatisfied needs for love. Transference reactions were originally considered to be the displacement of repressed fantasies from childhood onto the analyst, which could then manifest as resistance in

treatment (Zetzel, 1956/1990). It was understandable for Freud that his patients should relate to their analysts in manners consistent with their relations with parents in childhood. While understandable, Freud generally thought that transference reactions were "the most powerful" form of resistance that patients unconsciously employed against treatment progress and were therefore in need of examination. Freud also noted, however, that transference reactions could be helpful for the therapeutic work if a patient were to transfer feelings of affection towards the analyst, rather than projecting feelings of lust or hostility (Freud, 1912).

Freud (1912) thus divided transference reactions into two broad categories: negative and positive. The negative transference were reactions that were characterized as explicitly impeding treatment, such as projected feelings of hostility stemming from a relationship with a parent onto the person of the analyst. In contrast, the positive transference reactions potentially could be facilitative of the work, depending on their form. While Freud believed that a positive transference based on repressed erotic desires could be just as likely to produce resistance as a negative transference, a positive transference based on conscious feelings of "sympathy, friendship, trust, and the like" could be a principal element for success in treatment: it "is the vehicle of success in psycho-analysis exactly as it is in other methods of treatment" (Freud, 1912, p. 33). Freud goes further: "Where the capacity for transference has become essentially limited to a negative one…there ceases to be any possibility of influence or cure" (Freud, 1912, p. 34). It is this conceptualization of positive transference reactions that underlies contemporary alliance writings.

Safran and Muran (2000) suggest that analytic writings on the therapeutic relationship then branched into two paths. One track began with Ferenczi and extended into the work of object relations theorists and interpersonal analysts. Ferenczi (1932/1988; as cited in Safran and Muran, 2000) contended that patients needed to actively re-live past relationship problems within the therapeutic relationship, rather than simply recalling such problems in the abstract. Ferenczi also emphasized the analyst's personality and tendencies as a real person, and not simply a vehicle for patients' projections, as impacting transference enactments.

The second line of thought on the relationship developed through the ego psychologists who emphasized and re-focused attention on the personal components in the therapeutic relationship, over and above transference enactments. This was an attempt to highlight the interactions between analyst and patient as a focal concern in therapy (Safran & Muran, 2000).

Safran and Muran (2000) then trace the development of the alliance through Sterba (1934, 1940/1990) to Zetzel (1956/1990). In her work, Zetzel began to develop the modern notion of the alliance. She differentiated 'transference neurosis' from the alliance in the following way:

> A differentiation is made between transference as therapeutic alliance and the transference neurosis, which, on the whole, is considered a manifestation of resistance. Effective analysis depends on a sound therapeutic alliance. (Zetzel, 1956/1990, p. 138)

Zetzel emphasized that the alliance was rooted in early developmental experiences that set the stage for whether patients were able to form trusting connections with others. When

lacking, Zetzel argued that the analyst would need to provide a supportive environment to help develop an alliance. Zetzel used the metaphor of maternal nurturing on the part of the analyst in her conceptualizations of the alliance (Safran and Muran, 2000).

Greenson (1965/1990, 1967, 1971) furthered the work of ego psychologists in his division of the therapeutic relationship into the transference, the alliance, and the 'real' relationship. Greenson's work highlighted the importance of the alliance as an integral focus in analytic treatments in its own right; if a positive alliance did not develop in therapy, this was considered as a failure in technique on the part of the therapist. Greenson (1965/1990) underscored this idea:

> For successful psychoanalytic treatment a patient must be able to develop a full-blown transference neurosis and also to establish and maintain a reliable working alliance. The working alliance deserves to be recognized as a full and equal partner in the patient-therapist relationship. (p. 153)

Greenson (1967, 1971) later extended the demarcation of the therapeutic relationship by describing the alliance as based on the 'real' relationship between client and therapist. The real relationship for Greenson was the aspect of the relationship premised on the personal connection between therapist and client comprised of trust, respect, and care for each other (as cited in Safran & Muran, 2000). Greenson believed that the alliance was founded on this real relationship, rather than being a product of the transference. In addition to delineating elements of the therapeutic relationship, Greenson highlighted the high degree of overlap among those elements. As he noted, "There is no transference reaction, no matter how fantastic, without a germ of truth, and there is no realistic relationship without

some trace of transference fantasy" (Greenson, 1967, p. 219; as cited in Gelso, Perez-Rojas, & Marmarosh, 2014, p. 124).

Bordin's tripartite model. A key development in the history of the alliance was a seminal paper by Bordin (1979). In this paper, Bordin presented a pan-theoretical model of the alliance in which he suggested that the psychodynamic concept of the alliance was applicable for all modes of therapy (and many other change-focused relationships, such as that of student and teacher). He described the alliance as a special kind of collaborative relationship that was comprised of three key elements: the development of an emotional bond in the therapeutic dyad, collaborative agreement on the goals to which treatment is working, and the collaborative engagement in completing the tasks relevant for the mode of therapy being undertaken. In light of the importance of Bordin's (1979) model for the present study, each of these elements is described in further detail below.

*Bonds***.** The quality of the attachment and trust between client and therapist was one of the focal elements in Bordin's model. He noted that the strength and kind of relational bond developed in therapy would manifest differently depending on the therapeutic modality and the types of tasks asked of the client, but that some degree of real emotional attachment, regardless of therapeutic modality, was a requirement for successful treatment.

*Goals***.** The identification of clients' goals for treatment, and the subsequent agreement on those goals between clients and therapists, was a second key facet in Bordin's alliance model. A fundamental goal that Bordin identified as central for therapeutic work was an agreement that client difficulties were, to a large degree, a function of the clients' thoughts, feelings, and behaviours, and thus in need of modification

to help provide relief to the client. Bordin noted that more specific goals varied by therapeutic modality, but, regardless of the therapeutic orientation, agreement on the particular goals of the modality was highlighted as important for success in treatment.

Tasks. The engagement of clients in the therapeutic tasks asked of them was the final factor in Bordin's alliance model. Bordin again highlighted the notion that regardless of what the particular tasks consist (e.g., free association in psychoanalysis or behavioural tracking in behavioural therapies), clients' agreement to and engagement in completing such tasks were posited as essential.

Summary. Bordin saw that the alliance "between the person who seeks change and the one who offers to be a change agent is one of the keys, if not *the* key, to the change process" (Bordin, 1979, p. 252; emphasis in original). Bordin's aim in his seminal article was to help to stimulate research into the identification of effective therapy factors exemplified across therapeutic modalities, a search more recently referred to as identifying the 'common factors' in psychotherapy. In this vein, Bordin called for further research to investigate the applicability of his alliance formulation in all psychotherapies. He noted:

> The newness of the idea that the strength of collaboration between patient and therapist may have more to do with the effectiveness of the therapy than the particular methods chosen ensures that there will be little direct evidence available to test the proposition. (Bordin, 1979, p. 255)

It was in response to this call, among others, and the ongoing search for common factors in psychotherapy that a variety of measures designed to examine the strength of the alliance were developed.

Measuring the alliance. A number of quantitative measures have been developed to capture client, therapist and/or observer perceptions of the quality of the alliance across treatments. In the sample of studies collected for the meta-analysis published by Horvath et al. (2011), more than 30 distinct alliance measures were identified, such as the Working Alliance Inventory (WAI; Horvath & Greenberg 1989), the California Psychotherapy Alliance Scale (CALPAS; Marmar, Weiss & Gaston, 1989; Gaston, 1991), the Penn Helping Alliance Questionnaire (HAq; Alexander & Luborsky, 1987), and the Vanderbilt Psychotherapy Process Scale (VPPS; O'Malley, Suh & Strupp, 1983); all of the above have several variant forms either designed prior to or succeeding from those listed above. These questionnaires are rooted in related, albeit distinct, conceptualizations of the alliance; thus each measures the construct in slightly different ways (Horvath & Bedi, 2002). This is demonstrated in the finding that total scores from WAI, CALPAS and HAq correlate strongly with one another: $.75 < r < .85$ (Hatcher & Barends, 1996).

In their meta-analysis of 201 published and unpublished alliance reports, Horvath et al. (2011) reported that a version of the WAI was used in 39.8% of the identified studies (80/201). The WAI represented the most frequently used measure in their sample and was utilized roughly twice as often as the next most frequently used measures. The HAq (15.4%; 31/201) and the CALPAS (13.9%; 28/201) were the second and third most frequently used measures, respectively. In light of this predominance of the WAI in the research literature, it was chosen as the measure on which to focus in the present work. As such, the history of its development will be described in further detail below.

Working Alliance Inventory. The WAI (Horvath, 1981; Horvath & Greenberg, 1989) was grounded in Bordin's pan-theoretical model and was designed by Horvath and

Greenberg to measure the strength of the working alliance across all therapeutic orientations and offer both empirical and clinical value to those who used it. Following Bordin's (1979) model, the original WAI (Horvath & Greenberg, 1989) was designed as a 36-item self-report scale that measured the alliance across three subscales: Bond, Goals, and Tasks. On the WAI and its successors, respondents are presented with a series of items in the form of statements (e.g., 'My therapist and I agree on what is important for me to work on') and are asked to rate the frequency with which they agree with each statement on a Likert-type scale. In the original WAI (Horvath & Greenberg, 1989), this Likert scale ranged from 1-7; a rating of '1' on the scale meant that the statement was 'Never' true for a respondent, and a rating of '7' meant the statement was 'Always' true. It was originally designed with client- and therapist-rated versions, and later expanded to include an observer-rated version.

Horvath (1981) and Horvath and Greenberg (1989) outlined the steps through which they developed and validated the WAI. Potential items were first generated based on content analyses of Bordin's writings focusing on each of the three dimensions in his alliance model. This initial pool of items was then subjected to construct validation when it was submitted to seven alliance 'experts' to assess how appropriate each item was for measuring a facet of the alliance. These experts were asked to do two things: rate each prospective item as to how relevant it was for measuring the alliance (rated on a 5-point Likert scale) and identify which of Bordin's three components of the alliance each item exemplified. The initial pool was then condensed and sent to twenty-one practicing psychologists, who were not experts in the alliance, who then rated the remaining items using the same criteria as the expert raters. Relevant items were then selected and retained

based on the uniqueness of the underlying concept measured (e.g., expressing trust) and the degree to which the raters agreed that each item was related to its underlying alliance domain.

This initial version of the WAI was then pilot tested with a sample of graduate student therapists and their clients. WAI ratings at this stage underwent item analyses that examined the correlations among item scores, items' subscale scores, and total scale scores; ideally, item scores would correlate strongly within their respective subscales and moderately with total scores. Items that evinced low correlations or deviations from this ideal pattern were examined in greater detail. These analyses resulted in several small item-level grammatical changes and significant revisions to two of the items on the Tasks subscale. Reliability estimates for item homogeneity for the subscales ranged from .85-.88 and .68-.87 for the client and therapist versions, respectively. Reliability estimates for the entire measure were .93 and .87 for client and therapist versions, respectively. In their review of the history of alliance measures, Elvins and Green (2008) identified the WAI as a scale that has good validity and reliability support.

Tracey and Kokotovic (1989) later undertook a confirmatory factor analysis of the original WAI. They compared several possible factor models, finding the most support for a bilevel structure in which the WAI was ascertained primarily to assess a general alliance concept, as well as secondarily to measure the three distinct subscales (Bond, Goals, Tasks). In undertaking their analyses, Tracey and Kokotovic selected four items from each subscale that were the highest loading on each factor and proposed a new short-form of the WAI. They then examined this shortened instrument and determined that its item and scale

properties were acceptable. Thus Tracey and Kokotovic created the WAI-Short Form (WAI-S).

Hatcher and Gillaspy (2006) identified concerns with Tracey and Kokotovic's (1989) factor analysis, including a small sample size, participants' schedule of WAI completion (i.e., completing the WAI after the first session of therapy only), and outdated fit indices that were no longer considered to be within an acceptable range. In light of these concerns, Hatcher and Gillaspy sought to undertake further analyses in two samples to gauge the original WAI's factor structure and, ultimately, to offer a modified short-form of the WAI. They found that the models' fits were outside acceptable limits, as well as finding a high correlation between Goals and Tasks subscales (r = .99). Given these findings, Hatcher and Gillaspy then undertook an exploratory factor analysis in one of the samples to identify items that could form an alternate short-form to allow for greater differentiation between Goals and Tasks subscales. After identifying such items, Hatcher and Gillaspy conducted a confirmatory analysis in the second sample. They also conducted item response theory analyses that indicated that clients were unable to discriminate effectively among all seven points on the Likert response scale; thus they proposed, tested, and accepted a modified 5-point rating scale. This new form, the WAI-SR, demonstrated improved psychometric properties over the original short form. It is this new form, the WAI-SR, on which the present study is based.

Working Alliance Inventory – Short Revised. In light of the centrality of the WAI-SR to the present work and its frequent use in clinical practice, the scale itself is reproduced verbatim below (see also Appendix A). First, the title and instructions at the

top of the measure are included for reference. Next, the items and a reproduction of the rating scale are presented.

Working Alliance Inventory – Short Revised (WAI-SR)

Instructions: Below is a list of statements and questions about experiences people might have with their therapy or therapist. Some items refer directly to your therapist with an underlined space – as you read the sentences, mentally insert the name of your therapist in place of _____ in the text. Think about your experience in therapy, and decide which category best describes your own experience.

IMPORTANT!!! Please take your time to consider each question carefully.

1) "As a result of these sessions I am clearer as to how I might be able to change"

1	2	3	4	5
Seldom	Sometimes	Fairly Often	Very Often	Always

2) "What I am doing in therapy gives me new ways of looking at my problem"

5	4	3	2	1
Always	Very Often	Fairly Often	Sometimes	Seldom

3) "I believe _____ likes me"

1	2	3	4	5
Seldom	Sometimes	Fairly Often	Very Often	Always

4) "_____ and I collaborate on setting goals for my therapy"

1	2	3	4	5
Seldom	Sometimes	Fairly Often	Very Often	Always

5) "_____ and I respect each other"

5	4	3	2	1
Always	Very Often	Fairly Often	Sometimes	Seldom

6) "_____ and I are working towards mutually agreed upon goals"

5	4	3	2	1
Always	Very Often	Fairly Often	Sometimes	Seldom

7) "I feel that _____ appreciates me"

1	2	3	4	5
Seldom	Sometimes	Fairly Often	Very Often	Always

8) "_____ and I agree on what is important for me to work on"

5	4	3	2	1
Always	Very Often	Fairly Often	Sometimes	Seldom

9) "I feel that _____ cares about me even when I do things he/she does not approve of"

1	2	3	4	5
Seldom	Sometimes	Fairly Often	Very Often	Always

10) "I feel the things that I do in therapy will help me to accomplish the changes that I want"

5	4	3	2	1
Always	Very Often	Fairly Often	Sometimes	Seldom

11) "_____ and I have established a good understanding of the kind of changes that would be good for me"

5	4	3	2	1
Always	Very Often	Fairly Often	Sometimes	Seldom

12) "I believe the way in which we are working on my problem is correct"

1	2	3	4	5
Seldom	Sometimes	Fairly Often	Very Often	Always

Research on the alliance. What follows is a selective review of the research literature about the alliance and alliance ratings. Considering the majority of studies described below incorporated a number of alliance measures, such as in meta-analyses, this review is not focused on findings from studies solely utilizing the WAI or its successors. Given the prominence of the WAI in the literature, the majority of the studies described

below utilized the WAI in some capacity. The alliance measures used within each study are noted to provide greater context.

Meta-analyses based on quantitative ratings of the alliance have consistently shown that the alliance is positively related to therapeutic outcome (Horvath et al., 2011; Martin et al., 2000). In line with the findings from previous meta-analyses, Horvath and colleagues (2011) found an aggregated effect size of the relationship between alliance and outcome of .28. The aggregated statistic was derived by averaging across alliance measures, alliance rater, time of assessment, outcome measure, type of treatment, and publication source. The specific correlation between WAI ratings and outcome was also .28 with correlations between alliance and outcome on other measures ranging from .23 (CALPAS) to .29 (HAq and VPPS). Client and therapist ratings of the alliance were also shown to hold differing predictive power as assessed across the above alliance measures; clients' alliance ratings were somewhat more predictive of therapeutic outcome than were those of therapists with demonstrated correlations of .28 and .20 for clients and therapists, respectively (Horvath et al., 2011). While these effect sizes are small, Horvath and Bedi (2002) suggest that small effects are still meaningful ones in the complex arena of psychotherapy research, especially in light of the consistency of this statistical relationship across studies and the number of variables at work in psychotherapeutic treatments.

Certain client factors have been shown to influence the formation of the alliance. In one study (Patterson, Uhlin & Anderson, 2008), clients' pre-treatment expectations concerning their 'personal commitment' to therapy (i.e., their commitment to the therapeutic work and to using the therapeutic relationship as practice for non-therapeutic relationships) were related with the Bond, Goals, and Tasks sub-scales on the WAI-SR after

the third session. The authors interpret this finding to mean that clients who expect to take responsibility and commit to the work in therapy are more able to establish positive, collaborative relationships with their therapists. In another study (Iacoviello et al., 2007), clients rated their alliances, as measured on the WAI, as stronger if they were matched to treatments (pharmacotherapy or psychotherapy) for which they had indicated a preference prior to commencing treatment. The alliance has further been shown to partially mediate the relationship between pre-treatment expectations and therapeutic outcome in short-term individual therapy (Joyce, Ogrodniczuk, Piper & McCallum, 2003). Joyce and colleagues (2003) did not use an established alliance measure in their study, but rather asked clients six questions that focused on alliance-related content (e.g., whether the client felt understood by the therapist).

Bachelor, Meunier, Laverdiere and Gamache (2010) found that clients' attachment styles were related to alliance ratings as assessed on the WAI (as well as HAq and CALPAS). Clients who were securely attached were more likely to provide higher ratings of their alliances. Other client factors that may affect the formation of the alliance include clients' interpersonal schemes (Beretta et al., 2005) and their readiness for change (Prochaska & Norcross, 2002).

Importantly, clients' and therapists' ratings of the alliance have been shown to diverge (Tryon, Blackwell & Hammel, 2007). A meta-analysis by Tryon and colleagues (2007) found that the correlation between clients' and therapists' alliance ratings was only moderate (r = .36) across several alliance measures, including the WAI. Clients' ratings were also found to be significantly higher than those of therapists in the same dyad (d = .63). Of the 53 studies sampled by Tryon and colleagues, 33 utilized the original long-form

WAI (Horvath & Greenberg, 1989) or the WAI-S (Tracey & Kokotovic, 1989). These two measures in particular demonstrated differences between client and therapist ratings of *d* = .65 and *d* = .78, respectively. Using the WAI, Fitzpatrick, Iwakabe and Stalikas (2005) found that differences between client and therapist ratings did not change significantly over the course of therapy.

Tryon, Blackwell and Hammel (2008) have found that clients and therapists tend not to use the full range of response options when providing ratings on the alliance. In their meta-analysis, Tryon and colleagues (2008) found that clients and therapists tended to use only the upper 20% and 30%, respectively, of ratings on alliance measures. While the sources of this tendency were unknown, the authors suggested that it may be due to a social desirability bias, whereby clients and therapists, consciously or otherwise, may have wanted to make themselves, their therapists/clients, or their courses of treatment appear in a more favourable light. In this meta-analysis, studies reporting use of the WAI predominated, accounting for 60% of those included (using either the original WAI or the WAI-S). The statistic of interest in these analyses was the mean alliance rating across participants in each study as a function of the total points available on each rating scale utilized (measured after the third session). For example, Tryon et al. looked at studies using WAI and WAI-S ratings, which are both scales that use a 7-point Likert scale. Tryon and colleagues were interested in determining the average strength of clients' ratings relative to that 7-point scale. They found that clients' mean percentage of the maximum rating was 81% for the WAI and 82% for the WAI-S, meaning that clients' average ratings on the 7-point Likert scale were roughly 5.7/7. These authors suggested that such high ratings were the result of two possibilities: 1) These ratings were accurate reflections of strong, early-

phase alliances; however, they question this explanation by highlighting findings from studies about premature terminators who also provide high, albeit somewhat lower, alliance ratings (76% of maximum as opposed to 82%) across all measures; or 2) Clients often answered with the intention of responding in an artificially elevated and socially desirable fashion. The finding from Tryon et al. (2008) is consistent, albeit in a somewhat more extreme fashion, with Hatcher and Gillaspy (2006) who found that raters had difficulties differentiating among the lowest points on the original 7-point Likert scale as found on the WAI-S (Tracey & Kokotovic, 1989).

Samstag et al. (2008) conducted analyses examining the predictive impact that three working alliance ratings had on outcome status and premature termination; this study included the WAI. These authors utilized a sample comprised of therapeutic dyads representing five different therapeutic orientations, including variants of psychodynamic, cognitive-behavioural, and supportive treatment approaches. Analyses differentiated and examined three subgroups of clients (16 dyads per group): 1) Premature termination (clients who dropped out within the first third of the planned 30 sessions of treatment); 2) Good outcome (i.e., high symptom change); and 3) Poor outcome (i.e., low or worsening symptom change). Therapeutic outcome status was determined using reliable change scores (Jacobson & Truax, 1991). Samstag and colleagues (2008) found that the group of premature terminators provided significantly lower alliance ratings as compared to both groups of treatment completers. Early alliance ratings were somewhat lower in the poor outcome group as compared to the good outcome group, but these differences were not statistically significant. In undertaking qualitative analyses, the authors also found that poor outcome clients, not premature terminators, evinced the highest degree of

interpersonal hostility with therapists. These findings were interpreted to indicate that alliance ratings could be useful in helping clinicians identify clients at risk for early dropout, but that other therapeutic factors, such as managing hostility, were integral for improving outcomes for those who completed treatment.

Xu and Tracey (2015) outlined two schools of thought within the literature about the alliance's relationship with treatment outcome. They dubbed the first literature the 'relationship as strategy' school. This group posited the alliance as a causal predictor of treatment outcome, whereby the quality of the alliance influenced the degree of symptom change over the course of treatment, and a number of research studies have in turn provided further empirical support for this model (e.g., Horvath & Symonds, 1991; Horvath et al., 2011). Xu and Tracey described the second model as the 'relationship as outcome' model, in which symptom severity and change predicted the strength of the alliance. Integrating these two models, Xu and Tracey posited a reciprocal model in which symptom change and the alliance influence and predict each other over sessions. Using clients' ratings on the WAI-S, Xu and Tracey found that changes in the quality of the alliance predicted subsequent symptom change (controlling for earlier symptom improvement) while changes in symptom severity also predicted later alliance quality (controlling for prior alliance ratings). In short, Xu and Tracey (2015) argued that the quality of the alliance and symptom load interact and enhance each other over the course of treatment. Falkenstrom, Granstrom and Holmqvist (2014) presented similar findings in their work, further supporting the reciprocal relationship between symptom load and the alliance.

Hatcher, Barends, Hansell and Gutfreund (1995) conducted work assessing differences between alliance measures. They analyzed clients' and therapists' total scores

on the WAI, CALPAS and HAq, and tested a model of fit with one shared and two unique factors (one for each group) using confirmatory factor analysis. They found that the emphasis on helpfulness as a factor in the HAq best represented both clients' unique views, as well as therapists' and clients' shared views of the alliance, whereas the WAI's focus on tasks and goals of therapy was more representative of therapists' views alone. The factor loadings regarding the CALPAS fell between those of the WAI and HAq. These authors concluded that perceived helpfulness and a focus on goals and tasks in therapy represented different aspects of clients' perspectives of the alliance. They also argued that clients and therapists held different implicit theories of the alliance, despite each alliance measure's singular theoretical foundation, and suggested that further investigation into implicit rater attitudes was necessary.

Hatcher and Barends (1996) extended their findings from Hatcher et al. (1995) by applying exploratory factor analysis to clients' ratings to help understand clients' perspectives further. This work again focused on the WAI, CALPAS and HAq and analyzed clients' ratings on these scales to ascertain whether common, unarticulated factors cut across responses on all three measures. Once factors were derived and correlated with an outcome measure, the items with the highest loadings on each factor were given to independent raters to provide brief descriptions. The factor that the raters described as 'Confident Collaboration' formed the 'vital core' of the alliance from the client perspective, according to these authors; this factor had the strongest association with outcome (over and above the general factor). Hatcher and Barends (1996) argued that 'Confident Collaboration' captured clients' confidence and commitment to helpful therapeutic work that was perceived as having positive treatment momentum and purpose. The authors

argued that this factor was not an aspect of the theoretical foundation of any of the three measures analyzed and suggested that their results "point to the need for substantial revision of alliance measures" (p. 1334).

More recently, Bachelor (2013) undertook an exploratory factor analysis using data from clients and therapists provided on the WAI-S, CALPAS and HAq in which she conducted analyses within and across each measure. When looking at the measures in isolation, she found that the factors derived from her analyses showed low correspondence with the theoretical constructs underlying each scale. She also noted that clients' implicit conceptualizations not only differed from theory, but they differed from those of the therapists in her study as well. When analyzing across measures, Bachelor found that clients conceptualized the alliance as being comprised of six constructs: the collaborative relationship with the therapist, the therapy's productiveness, a commitment to the process of therapy, the affective bond, non-disagreement on tasks and goals in therapy, and confidence in the progress realized from therapy. She concluded that while therapist perspectives on the alliance could be captured fairly well by the measures as they were, client views indicated the possible need for some modification of the existing scales.

Qualitative investigations of the alliance. Relative to quantitative investigations, comparatively little work has examined clients' qualitative, subjective perceptions of the alliance. In an early work utilizing phenomenological and content analyses, Bachelor (1995) showed that clients tended to view a positive alliance in types, as either nurturant (i.e., based on trust and therapist attitudes, such as non-judgment, respect, and expressed empathy) or insight-oriented (i.e., based on acquiring insight into difficulties). To a lesser degree, some clients viewed a positive alliance as the result of their active involvement in

therapy. She noted, however, that most clients endorsed therapist qualities, such as respect, non-judgment and careful listening, as being facilitative, regardless of their preferred overall type of alliance.

Fitzpatrick, Janzen, Chamodraka, Gamberg, and Blake (2009) found that when clients with self-described positive alliances were asked to think of a 'critical incident' that represented their experiences in therapy during a semi-structured interview, they noted several important therapist influences, including the facilitation of thinking or acting in a new way, the provision of emotional support, the communication of understanding, the meeting of unexpressed needs, and a nonjudgmental style. Fitzpatrick and colleagues found that clients tended to view these events as evidence that their therapists cared about them.

Bedi, Davis and Williams (2005) conducted an exploratory study using semi-structured interviews to ascertain what factors clients believed to be important in the formation of a strong positive alliance. The five most common variables that participants viewed as being particularly important for positive alliance formation were the technical activities of the therapists, such as teaching a grounding technique (though excluding active listening), therapists' nonverbal communication, therapists' active listening skills, therapists' emphasis on client choice and control in therapy, and the therapy environment (i.e., the accoutrements of therapists' offices). These authors noted that some of the variables identified by clients, namely, technical activity and the therapy setting, are seldom researched in the context of alliance formation. This finding led the authors to argue that client- and researcher-driven perspectives of positive alliance formation may differ with respect to which variables are determined to be of interest. As with client and

therapist divergence in terms of alliance ratings, these findings suggest that there may be divergence between client and researcher conceptualizations of the alliance.

In light of findings of differences between clients' and both researchers' and therapists' conceptions of the alliance and the relationship between clients' ratings and outcome status, there was a clear need for further research to better understand clients' perspectives on this key psychotherapy variable. As noted earlier, the WAI (Horvath & Greenberg, 1989) and the revised editions based on it (Hatcher & Gillaspy, 2006; Tracey & Kokotovic, 1989) are the most common measures used to assess the alliance (Horvath et al., 2011), and as such, are a key focus for research seeking to enhance our knowledge of the alliance and what clients convey to us with their alliance ratings. In order to further understand the WAI and clients' approaches to it, it is important to have an understanding of respondents' approaches to self-report rating scales in general.

Evaluating Respondent Approaches to Self- Report Rating Scales

Self-report measures in general (Schwarz, 1999), and rating scales in particular (Rosenbaum & Valsiner, 2011), have been argued to yield potentially problematic data with regard to psychological processes. They have been criticized on the grounds that responses gleaned from such measures are not necessarily indicative of respondents' internal processes or behaviours and should not be thought of as objective assessments (Schwarz, 1999). Schwarz (1999) has argued that the self-report measure, a mainstay in both basic and applied psychological research, can result in divergent conceptualizations between researchers and respondents of the questions and issues under study. Such variables as question formatting, question wording, and the context in which respondents encounter each question may influence how participants respond to self-report items.

These influences are at work when respondents complete any type of self-report measure, including rating scales.

Self-report measures have also been criticized for propagating researchers' worldviews and biases, resulting in limited insight into respondents' lived experiences (Galasinski, 2008; Galasinski & Kozlowska, 2010). Smith (2008) argued that self-report measures constrain responses to the point of pre-determining an item's outcome. Smith investigated the communicative actions of respondents when completing a pain experience questionnaire that had been delivered through the mail. She found that a number of respondents wrote in the margins of the measure to elaborate or qualify their item responses because, Smith argued, respondents found the item response options to be insufficient.

Ratings scales in particular have been suggested to imprecisely quantify experiences that inherently exceed quantification. Rosenbaum and Valsiner (2011) argue that researchers who analyze data derived from rating scales make several problematic assumptions about respondents' answers: 1) Respondents comprehend and have direct access to the meanings of the poles that anchor the ends of the rating scales; 2) They accept and can distinguish between the continuous points on the scales; and 3) Their understandings of each item on these scales are sufficiently similar to enable the aggregation of data within and between multiple respondents into a single sample.

The standardization of surveys and survey procedures has been a primary focus for researchers. Implicit in standardization, though too-often overlooked, is the assumption that all respondents understand the questions in the same manner and that they are able to respond accordingly (Collins, 2003; Rosenbaum & Valsiner, 2011). Since respondents must

ascertain both the literal meaning of each item, as well as make inferences as to the item's pragmatic meaning, convergence across participants cannot be assumed and should be demonstrated empirically over the course of measure development (Schwarz, 1999, 2007). This assumption that all respondents understand items in the same way is often uninvestigated in the standardization and pilot phases of survey design, even though the validity and reliability of the results gleaned from surveys rests fundamentally on the common understanding across respondents. Employing cognitive interview methods to test questionnaire items can help to establish that respondents understand all questions in a similar and consistent way and that their understandings of questions are in line with those the researchers intended (Collins, 2003). Cognitive interview methods were utilized in the present study and are described in greater detail below.

Survey response theories. Tourangeau and Rasinski (1988) have argued that questionnaire respondents need to undertake and resolve four demands in order to answer a questionnaire item: 1) They must comprehend that item; 2) They must retrieve the relevant information being queried; 3) They must evaluate that information and estimate a response; and 4) They must actually provide a response (Willis, 2004). Called the Tourangeau-Rasinski model (Tourangeau & Rasinski, 1988; Tourangeau, Ripps & Rasinski, 2000), these demands, conceptualized as stages in a survey completion process, are referred to as question comprehension, information retrieval, judgment and estimation, and response (Grant, Rohr & Grant, 2012). Collins (2003) argued that these stages were not completed in a linear fashion and involved oscillation and interaction among stages. Rather than conceiving of them in purely linear terms, they can be thought of as four semi-distinct demands that respondents necessarily meet when answering an item on a scale. In

order to better appreciate the processes that underlie survey responding and the requirements on researchers when ensuring the intelligibility and appropriateness of survey questions, I will briefly summarize each stage of the process in turn, following summaries found in Collins (2003) and Grant and colleagues (2012).

Stage 1: Question comprehension. This stage entails that all respondents understand questions on the survey consistently and in the manner in which the researchers intended. Respondents need to understand items in a similar manner to other respondents so that data across respondents can be aggregated. Additionally, if respondents understand a question in a manner that is inconsistent with that intended by the researchers, conclusions drawn by researchers from those responses may be inaccurate. Using a relevant example from the WAI-SR, clients completing the WAI-SR will encounter one item that reads "I feel that _____ (my therapist) appreciates me." As noted above, clients completing the WAI-SR are asked to assess the frequency with which they agree with a given statement, such as the item just noted, on a 1- ('Seldom') to 5- ('Always') point Likert-type scale. For this item at this comprehension stage, respondents will need to understand the words used in and the intention of this statement in a similar fashion as other respondents in order to reliably aggregate ratings from this item. Clients will generally also need to comprehend the words and intention in a similar fashion to those of researchers or clinicians; otherwise misinterpretations on the part of those receiving clients' responses can arise due to inconsistent understandings.

Stage 2: Information retrieval. Once the question has been understood, respondents must then access and retrieve the relevant information that is being solicited. In our earlier example, clients presented with the WAI-SR item "I feel that _____

appreciates me" must identify relevant experiences (memories, feelings, etc.) that are pertinent to the question, and, by extension, must identify which experiences are not relevant and should thus be disregarded. In short, if clients have just read "I feel that _____ appreciates me" and have comprehended the question, they must now identify and retrieve the information that is relevant for their answer, such as memories of when the therapist demonstrated an attitude that was interpreted as appreciative.

Stage 3: Judgment and estimation. This stage involves how respondents formulate answers for questions and is applicable during all phases of the item response process (i.e., during comprehension, retrieval, and response). This is the phase in which respondents ascertain whether they understand the literal and pragmatic facets of the question, whether and how it applies to them, whether they have the relevant information needed to answer the question, and so on. For example, clients assessing the frequency of their agreement with "I feel that _____ appreciates me" will judge whether they understand the question, whether relevant experiences are accessible to memory, whether a quantitative rating can be applied to those experiences, and so on.

Stage 4: Response. Once respondents have made a judgment as to how to respond to a particular item, they often then need to fit their desired response into the pre-set response options. On the WAI-SR, respondents must fit their own experiences and judgments of frequency into one of the pre-defined categories offered to them on the 5-point Likert scale. The descriptors of the response categories (i.e., the qualitative labels applied to each point on a rating scale, e.g. 1 = Seldom) can impact how respondents comprehend, judge and respond to each item. An additional aspect of this phase may be

respondents' tendencies to modify responses to ensure positive self-presentation and conform to social norms; thus they may be circumspect when providing their answers.

Understanding survey responses in the theoretical terms and phases that are summarized above enables researchers to highlight possible problems with survey items with greater specificity, to better understand and interpret respondents' answers, and to specifically target any item modifications as needed (Collins, 2003).

Research on survey methods. The roots of cognitive survey interviewing methods can be found in the Cognitive Aspects of Survey Methodology conference held in 1984 (Beatty & Willis, 2007). Generally, cognitive interviews are undertaken for the purposes of engaging in the initial testing and modification/improvement of surveys, but they have also been seen as a valuable method with which to continuously evaluate existing instruments (Karabenick et al., 2007). When modification is not the primary focus or is not possible (e.g., in the case of an existing measure), the findings from cognitive interviews can be used to aid in the interpretation of respondents' answers (Adair et al., 2011).

The question of how to most effectively undertake such investigations, however, is still one of debate. There are two principal methods with which such interviews are conducted: think-aloud reporting and verbal probing (Beatty & Willis, 2007; Collins, 2003). A think-aloud procedure is one in which respondents are asked to narrate their processes (e.g., thoughts, memories, feelings) when completing a measure. Interviewers initially do little in these interviews other than to introduce the task and to direct respondents to narrate their experience. One fundamental assumption of this method is that respondents can retrieve and accurately report the relevant information, such as real-time meta-cognitive reflections of their comprehension of particular questions. This method also

assumes that the act of reporting does not change the reported activity itself (i.e., does not produce reactivity in respondents). It is also assumed that its lack of interviewer involvement results in less interviewer influence over responses than more structured interviews, and thus provides a relatively unfiltered view into respondents' experiences (Priede & Farrall, 2011).

Verbal probing is a more structured type of cognitive interview in which interviewers explicitly ask questions of respondents, again with the intention of understanding respondents' processes while engaged in answering self-report questions (Beatty & Willis, 2007). These interviews generally ask respondents to explain how they comprehend each item in a survey, what they would provide as answers, and why/how they came up with their answers (Karabenick et al., 2007).

In a study that explored the effectiveness of both interview procedures by piloting a survey that assessed the public's confidence in the U.K.'s justice system, Priede and Farrall (2011) found that verbal probing interviews identified approximately 1.5 times as many item problems, though they also found that this method introduced information into respondents' accounts, based on interviewers' questions, that was not contained in respondents' original answers. They noted that verbal probing did allow for greater depth of understanding of a particular issue, whereas no such elaboration was possible during think-alouds. They concluded that an interview protocol that incorporates the openness of think-alouds with the querying flexibility of verbal probing may be the best method for survey evaluation.

Although both cognitive interview procedures have theoretical benefits and areas of difficulty, little empirical research has been conducted to compare the think-aloud and

verbal probing methods to assess the quality of the findings derived from each (Priede & Farrall, 2011). Guidelines for best practices regarding cognitive interview data collection and analysis are lacking and, given the absence of such research, in need of further study and delineation (Beatty & Willis, 2007). Importantly for the present study, neither interview method has been utilized to evaluate clients' responses to items on any version of the WAI.

Summary

The working alliance as a construct, and the measurement of it via questionnaires, represents a significant area of psychotherapy research within which numerous questions remain to be answered. Since the alliance is a concept in need of greater clarification and, perhaps, modification (Safran & Muran, 2006), an examination of the most widely used rating scale – the WAI - through which the construct is measured is a necessary next step. While many of the qualitative studies discussed above provide key insights into client perspectives on the alliance and WAI responses in particular, questions still remain as to how clients' qualitative perspectives are related to self-report quantitative ratings on the WAI. Since it is clients' ratings on the WAI that are the most predictive of therapeutic outcome (Horvath et al., 2011), improving our understanding of client ratings is essential for understanding the relationship between the alliance and therapeutic outcome. Following the meta-analysis from Horvath and colleagues (2011) in which approximately 40% of the identified studies utilized a version of the Working Alliance Inventory, a considerably higher percentage than any other measure, understanding client perspectives on this principal alliance measure is particularly important.

While Hatcher and Barends (1996) and Bachelor (2013) have conducted valuable investigations into client perspectives on rating their alliances from a quantitative perspective, no study to date has explored clients' understandings of their own ratings using qualitative methods. Several exploratory research questions addressing this gap in the research literature inform the present work: How do clients understand the WAI itself and the ratings they provide? What are they referencing when they provide their answers? How do clients' quantitative ratings and their qualitative understandings of those ratings converge or diverge? In short, without knowing *how* clients rate their alliances on measures, it is difficult to know *what* it is that they are rating. The content of their ratings, the *what*, has been the focus of numerous studies in the literature, whereas the process of their ratings, the *how*, has been relatively ignored.

Accordingly, the present study targeted more specific item information regarding the intelligibility of the WAI as a measurement scale. Considering that accurate quantitative assessment is predicated on consistent understandings of questionnaires across participants, an item-specific focus was also needed. Karabenick and colleagues (2007) argued that cognitive interview methods, including think-aloud protocols and verbal probing, could be a fruitful group of qualitative approaches to employ both in the development and throughout the life of a scale. They suggested that evidence from these methods in particular could inform future theory development of the construct being measured, as well as aid the interpretation of results derived from the instrument.

The Present Study

The present study sought to clarify and provide insight into the processes through which clients provide their ratings on the WAI-SR. To do so, the study utilized qualitative

interviewing and analytic methods to explore clients' perspectives of WAI-SR items and the measure as a whole, as well as clients' rating processes engaged when providing ratings. Interviews were undertaken with clients, the majority of whom were undergraduate students receiving therapy through the York University Psychology Clinic, the York University Psychology Practicum Clinic, and a research trial assessing the effectiveness of Cognitive-Behavioural Therapy with an adjunctive element of Motivational Interviewing (Westra, Constantino, & Antony, 2016). Most interviews were completed within 24 hours of clients' final sessions with their therapists.

Specifically, this study included three phases. In the first phase, a semi-structured interview protocol and grounded theory analyses were employed to ascertain clients' general perspectives of the WAI; twelve interviews were completed for this phase. In the second phase, two cognitive interviewing methods were utilized and content analyses conducted to identify specific client concerns about each of the items on the WAI-SR; any concerns were then described qualitatively. Ten interviews were undertaken for this phase. The third phase comprised a triangulation of findings from the first and second phases and the presentation of an intensive single case interview analysis in the service of developing a comprehensive description of client perspectives. Each of these phases was designed to explore the research questions described below.

Exploratory research questions. ***Phase one: Of what are clients aware when completing the WAI-SR and on what do they base their ratings?*** Considering Hatcher and colleagues (1995), Hatcher and Barends (1996) and Bachelor's (2013) suggestions, derived from quantitative studies, that clients' perceptions of the alliance are not identical to those of researchers and clinicians, a qualitative investigation into clients' internal

experiences of the alliance, and how such experiences influence the ratings that they provide, is a necessary next step in understanding clients' ratings. Schwarz (1999) and Rosenbaum and Valsiner (2011) have argued that rating scales of psychological processes produce results that are not necessarily representative of respondents' internal experiences, and thus researchers' assumptions that the WAI-SR is measuring what they believe it to be measuring requires validation against clients' experiences. To address this question, a semi-structured interview protocol was employed to access clients' experiences of engaging with the WAI-SR. Twelve interviews were completed during this phase, with the majority taking place within 24 hours of participants' final treatment sessions. All participants were York University undergraduate students; eleven of these participants were receiving treatment through the York University Psychology Practicum Clinic while the twelfth was taking part in a research trial. Grounded theory analyses were undertaken to examine the findings of these interviews.

Phase two: What concerns do clients identify when completing the WAI-SR items? In light of Gilgun (2004) and Barroso and Sandelowski's (2001) calls for qualitative explorations of quantitative measures, as well as Karebenick et al.'s (2007) contention that think-aloud and verbal probing protocols are two methods with which to undertake such investigations, think-aloud and verbal probing interview procedures were utilized to explore this research question. Interviews focused on clients' comprehension and evaluation of WAI-SR items and response options to ascertain if clients experienced any concerns or difficulties when rating. A total of 10 interviews were completed for this phase, five of which utilized a think-aloud procedure while the remaining five employed a verbal probing protocol. All interviews occurred within 24 hours of participants' therapy

sessions, with seven of the ten interviews taking place after participants' final treatment sessions. Participants were recruited using convenience sampling from the York University Psychology Practicum Clinic and the York University Psychology Clinic. A pre-determined coding system based on the Tourangeau-Rasinski model for survey responding (described in Rothgeb, Willis & Forsyth, 2005) was employed to analyze results from both interview methods. Considering the widespread use of the WAI and its offspring in measuring the alliance (Horvath et al., 2011) and the absence of substantial qualitative checks on clients' interpretations of this measure, this question is an important one that has as yet gone unaddressed in the literature.

Phase three: How do clients understand, engage with, and respond to the WAI-SR overall? Miles and Huberman (1994) suggest that the integration of findings from the use of multiple methods can enhance the objectivity of a research endeavour. They argue that employing triangulation, a process through which a researcher can demonstrate that multiple sources of data agree with one another, can increase confidence in a study's results. Interview results from the three interview protocols were triangulated using a procedural approach. Following Meijer, Verloop and Beijaaard's (2002) description of the method of procedural triangulation integrated with those described in Farmer, Robinson, Elliott, and Eyles (2006), the results produced from research question one (i.e., a general theory of how clients engage with their relationships with their therapists, and how these are represented on the WAI-SR) were triangulated with the results from research question two (i.e., the identification of any salient item- or scale-specific problems with the WAI-SR) and supplemented with a single case analysis to provide a more global understanding of how clients understand and respond to the WAI-SR. It was anticipated that by utilizing

three distinct interview methods, as well as qualitative and content analyses, a deeper integration of research findings would emerge for the identification of multiple facets of clients' engagement with and perspectives of the WAI-SR.

Method

Overview

The first phase of this study employed a semi-structured interview protocol with 12 participants. Questions in this phase encouraged participants to reflect on their engagement with the measure and their therapists. Grounded theory analyses were conducted on these participant narratives. The second phase employed cognitive interview methods (think-aloud and verbal probing) with 10 participants in total, five with each protocol. These interviews focused on item-specific data, encouraging participants to reflect on their ratings of all items in succession. The Classification Coding Scheme (Rothgeb et al., 2005) was used to describe these interview data. The third phase integrated findings from both interview phases, assessing the degree of convergence of results from primary phases. The methods employed during each phase will be described in more detail below.

Considering that participant sample sources and recruitment were virtually identical in the first two primary phases, these procedures are described first. Phase-specific procedures regarding specific participant characteristics, interview protocols, and analytic methods are then described.

Procedure

The procedure for participant recruitment was the same for both phases of the project. Institutional ethics approval was acquired from York University's Ethics Review Board prior to beginning recruitment for the final phases. This researcher completed all participant interviews in both phases.

Twenty-two participants in total took part in interviews for this project: twelve in Phase One and 10 in Phase Two. For Phase One, 11 of the 12 interviews were completed following clients' final sessions of therapy (the one exception followed a client's penultimate session) to allow them the greatest ability to reflect on their experiences over therapy and with the scale. For Phase Two, seven of the ten interviews took place following clients' final sessions, while the remaining three interviews were with participants who were continuing in treatment. The majority of interviews in both phases occurred within 24 hours of participants' most recent session, with several exceptions described below.

Sample sources. Participants were recruited from two clinics located at York University, as well as an ongoing research study investigating the integration of CBT and Motivational Interviewing (MI) for Generalized Anxiety Disorder (GAD) that was affiliated with York and Ryerson universities. All participants provided informed consent to participate in this study (see Appendix B for consent form). Each sample source is described in further detail below.

York University Psychology Practicum Clinic. The first and principal source for participants in this study was the York University Psychology Practicum Clinic (YUPPC). All clinicians in this clinic are current graduate students in the Clinical Psychology program working under the supervision of faculty members; most clinicians are second-year Masters or first-year doctoral students. All clients of this clinic are current students, typically undergraduates, at York University who received psychology services as a part of their student health benefits; they thus received services without additional out-of-pocket charge.

YUPPC clients were initially screened by clinicians at York University's Counselling and Disability Services program for possible appropriateness for therapy with graduate-level therapists in the YUPPC. YUPPC supervisors then conducted a second screen of all prospective clients to ensure this appropriateness. Exclusion criteria for prospective clients included: 1) History of major trauma; 2) Active suicidality; 3) Psychosis; and 4) Eating disorders. All YUPPC clients completed process measures, including the WAI-SR, as a regular part of treatment. YUPPC clients typically completed the WAI-SR after every other session. Eighteen participants were recruited through the YUPPC (11 in Phase One, seven in Phase Two).

York University Psychology Clinic. The second source for participants was the York University Psychology Clinic (YUPC), a fee-for-service community mental health clinic associated with the Psychology Department at York University. The majority of clinicians within this clinic are current or former graduate students and faculty members in the Clinical Psychology program. Clients are comprised of York students and staff, as well as community members. Only some clients in the YUPC completed process measures, such as the WAI-SR, as a regular part of their treatment; those who did complete them tended to do so after every second session. Three participants were recruited from this source (one in Phase One and two in Phase Two).

***MI-CBT Trial*.** The final recruitment source was an ongoing research trial examining the effectiveness of utilizing MI prior to undertaking traditional CBT versus CBT alone in the treatment for severe GAD (Westra et al., 2016). This trial was run through York and Ryerson universities. All clients in this trial met DSM-IV-TR criteria for GAD of high severity (Penn State Worry Questionnaire score of 68 or higher) prior to inclusion. Clients in this

trial completed a series of process measures throughout treatment, one of which was the WAI-S (Tracey & Kokotovic, 1989). One participant was recruited from this source (MI-CBT treatment condition) and was interviewed during Phase One; she was the only participant who completed the older WAI-S for interview.

Recruitment. Specific recruitment procedures utilized for clients from each of these sources was similar, albeit slightly different. The following describes the recruitment procedures in general, which were identical across Phases One and Two.

1) Active clinical supervisors in the YUPC, YUPPC and a co-principal director of the GAD trial were consulted about allowing the researcher to contact supervised therapists to acquire consent to be involved in the study. Supervisors provided the researcher with the names of possible therapists.

2) Therapists were emailed information about the study and its voluntary nature. They were asked if they would provide consent to inform their clients about the study (where clinically appropriate). Therapists were notified that they would also be asked to complete a brief measure about their demographic and therapeutic information if one or more of their clients ultimately participated.

3) Therapists who responded and consented were sent information to provide to their clients about participation in this study. This package included: a) A client flyer, introducing the study and asking for client's consent to be contacted by the researcher for more information, as well as asking for their name and phone number; b) A script with which to notify clients of the flyer; and c) Specific procedural instructions for returning completed flyers, which varied slightly by recruitment source. In the majority of cases, therapists were asked to include this flyer in their clients' post-session questionnaire

packages. Differences among the clinic sources arose at this point, as the more junior Masters-level clinicians did not open and view their clients' post-session measures; these were reserved for supervisors.

4) The researcher contacted clients who completed flyers in the week following receipt of the flyers. The researcher explained the study in further detail, highlighted benefits and risks, and asked for verbal consent to schedule an interview. Interviews were scheduled to follow a session of therapy, ideally taking place after clients' final sessions when possible.

5) The researcher met participants after their scheduled session. Prior to beginning the interview, the researcher again reviewed relevant information, provided clients with an informed consent form (see Appendix B), and acquired informed consent to proceed. Clients were also asked to complete a brief demographic questionnaire at this time (see Appendix C).

6) Following interviews, clients' post-session measures were returned to therapists or supervisors, as relevant.

Measures

Demographics Questionnaire. Participants provided information about their age, gender, achieved education, York University status, marital status, length of treatment in current and previous courses of psychotherapy, and whether they were paying for their current courses of treatment. See Appendix C for this questionnaire.

Working Alliance Inventory – Short Form (Patient Version). This first short form (Tracey & Kokotovic, 1989) was derived from the original 36-item WAI (Horvath & Greenberg, 1989) described above. It is a 12-item scale on which ratings are made using a

7-point Likert-type scale ranging from 1 ("Never") to 7 ("Always"). Items cluster on three subscales in line with Bordin's (1979) tripartite alliance model: Bond, Tasks, and Goals. This measure has demonstrated adequate psychometric properties. Only one participant in the present study, WAI_08 in Phase One, completed this version of the WAI.

Working Alliance Inventory – Short Revised. This revised short form (Hatcher & Gillaspy, 2006) represents a more recent iteration of the original WAI (Horvath & Greenberg, 1989). Similar to the WAI-S, this is a 12-item scale that includes three, four-item subscales: Bond, Tasks, and Goals. Ratings are made on a 5-point Likert-type scale ranging from 1 ("Never") to 5 ("Always"). This version of the measure has demonstrated good psychometric properties. All participants with the exception of WAI_08 completed this version of the WAI in the current study. See Appendix A for this scale.

Phase One Method

The first phase of this project was conducted to address research question one; namely, this phase sought to understand what clients were aware of and to what they responded while completing the WAI-SR. This phase was comprised of semi-structured interviews completed with 12 clients in ongoing therapy and used the Ground Theory method to analyze interview data.

Participants. Participants for this phase were 12 clients who had been receiving ongoing psychotherapy; 11 of these clients received therapy through the YUPPC, while one participant was recruited from the MI-CBT for GAD research trial. All participants were undergraduate students at York University. Interviews for this phase were completed between August 2013 and April 2014.

The two inclusion criteria for participants in this phase were: 1) They completed the WAI-SR as a regular part of their work with their therapists; and 2) They had filled out the WAI-SR in the therapy session immediately preceding interviews. There were no specific exclusion criteria for the present study (other than participants having already been screened by supervising psychologists prior to entry into the YUPPC and the CBT trial).

Participant data for this phase can be seen in Table 1. Nine women and three men with an average age of 23.1 (range = 19-39) took part in interviews for this phase. One participant had received psychotherapy elsewhere for three months, but no other participant had received therapy previously. Mean WAI scores were computed for each participant across all sessions. The mean WAI score across the sample was 4.24, which appears to be consistent with sample averages reported in other studies utilizing the WAI-SR or WAI-S (e.g., Falkenstrom et al., 2014 (4.08 equivalent); Reese et al., 2013 (4.40); Tryon et al., 2008 (4.09 equivalent); Turner, Bryant-Waugh & Marshall, 2015 (4.33 equivalent)).

Therapists. There were 10 therapists seeing clients in this phase. Two therapists saw two participants each, while the remainder saw a single participant. All therapists were graduate trainees at York University; three were doctoral trainees while seven were Masters-level. Eight were women and two were men. Mean age for these therapists was 27.8 years (range = 26-31) and had a mean of 2.3 years (range = 1-7) therapy experience.

Interviews. The semi-structured interviews during Phase One lasted 40-90 minutes. All interviews were audio-recorded and transcribed. Eleven of the 12 interviews were completed following participants' final sessions in therapy, with the exception of one interview that was completed after a participant's penultimate session. Ten of the 12

Table 1
Phase One Participant Characteristics

Client ID	Age	Gender	Source	# Sessions[a]	Mean WAI
WAI_08	19	F	MI-CBT Trial	15	4.88[1]
WAI_11	19	F	YUPPC	17	4.33
WAI_12	22	M	YUPPC	17	4.04
WAI_13	31	F	YUPPC	13	4.73
WAI_14	20	M	YUPPC	17	3.68
WAI_15	19	F	YUPPC	15	4.44
WAI_16	39	F	YUPPC	13	4.15
WAI_18	21	F	YUPPC	15	4.00
WAI_19	23	F	YUPPC	17	4.31
WAI_20	22	M	YUPPC	15	3.90
WAI_21	23	F	YUPPC	17	3.53
WAI_22	19	F	YUPPC	10	4.92

[a] This column indicates the number of sessions each participant completed prior to interview.

interviews were completed within 24 hours of participants' most recent therapy sessions. However, two interviews were delayed for scheduling reasons, with a gap of two days in the first case and one week in the second.

The interview procedure utilized in this phase involved a semi-structured protocol that began with the interviewer inviting each participant with an open and general prompt, "Tell me what you were aware of today when completing the WAI." This open prompt allowed participants to provide reflections without being channeled by interviewer queries. The present interview procedure was designed as a narrative interview to encourage opportunities for client reflection where the interview begins more generally and becomes progressively more focused throughout. This process, and the incorporation of relatively unstructured questions in particular, allows the interviewer greater access to the meaning conveyed in participants' narratives (Corbin & Strauss, 2008; Groleau, Young

[1] As this participant completed the original WAI-S, which employed a 7-point scale, this mean WAI score has been converted into its 5-point equivalent for ease of comparison across participants.

& Kirmayer, 2006). After participants responded to this initial prompt, the interviewer attended to salient elements by offering summaries and further queries in a process of 'guided conversation' (Kvale, 1996). The flow of the interview and the content under scrutiny was co-constructed by participant and interviewer, and many of the interviewer's questions developed over the course of each interview (Corbin & Strauss, 2008). The researcher did not ask questions targeting specific items in this phase unless participants explicitly highlighted a particular item.

Based on the initial review of the literature, discussions with committee members, and the findings from pilot phase interviews, the interviewer inquired about specific areas of interest during this phase. In many cases, the interview process unfolded in such a way that participants provided information related to these areas of interest organically over the course of interviews, but the interviewer made explicit queries at the end of any interviews where these areas of interest had not been raised already; they were not targeted to specific items. These queries were the following:

1) What do you think the questionnaire is asking of you in general?

2) How well do you think this questionnaire captures your connection with your therapist?

3) How do you come up with your answers to the questions in general?

4) Have your ratings changed over time? If so, how?

5) Have you interpreted these items differently over time? If so, in what way?

6) Did factors outside your course of psychotherapy impact your ratings, such as your mood that day, environmental stress, dislike of the questionnaire, etc. If yes, how?

7) Do you think that filling out this questionnaire has affected your working relationship with your therapist in any way? If so, how?

8) How did you feel about being asked about your relationship with your therapist? How did you feel about then being asked to rate it?

9) Who saw the ratings you provided? Did you think about who would see them when completing the questionnaire? How did this impact the ratings you gave?

10) Are there any items you find particularly challenging or confusing?

11) If you are unsure of what a question is asking of you, what do you do?

12) If you are unsure of your answer to a question, what do you do?

13) How do you think this questionnaire was used throughout your therapy?

14) What differentiates, for example, a rating of a three from a four, or a four from a five in general?

15) What would you add to this questionnaire, if anything?

16) In general, what does a good client-therapist relationship look like? What are its features?

Analysis. The researcher undertook a Grounded Theory (GT; Glaser & Straus, 1967; Corbin & Strauss, 2008) analysis of participants' narratives acquired in these interviews. GT was chosen as the analytic method for this phase as it allows for the analyst to develop a thorough understanding of participants' experiences (Kagan, 2014) and is widely used in clinical and counselling psychology qualitative studies (Ponterotto, 2005). There are several variants of the GT method that have developed over time (Fassinger, 2005), but at the heart of all GT analyses is the intention towards generating a substantive and coherent theory of participants' accounts of their experiences in regards to a particular phenomenon

under study (Rennie, 2006). The theory produced through GT analyses is "grounded" in that it is rooted in the interview accounts that participants provide. The theory developed through GT analyses is constructed by coding and analyzing transcripts of participant accounts and comparing new transcripts with previously analyzed interviews until no new themes or categories are generated in the analysis. Codes are clustered into categories as individual accounts are analyzed. Once new categories cease to develop (i.e., once category saturation is reached), relationships connecting categories are highlighted to help create a theory about the phenomenon under investigation (Fassinger, 2005).

More specifically, GT analyses are undertaken in three stages (the exact stages depend on which GT variant one is utilizing): 1) Open coding, wherein transcripts are broken down into meaning units (concepts) and identified as properties that are subsequently grouped together to form more abstract categories of responses; 2) Axial coding, wherein initial categories are grouped together to form higher-order (key) categories in the service of noting the relationships between lower-level categories and concepts; and 3) Selective coding, in which a core category is formed to integrate and explain other categories in the study. At this stage, a theory is generated that accounts for participants' experiences. Throughout all phases, the researcher writes notes (memos) concerning decisions made and questions raised. This is done in the service of keeping a record of the development of ideas and assumptions during analysis (Corbin & Strauss, 2008; Fassinger, 2005).

Interviews were transcribed and analyzed chronologically, and an experienced GT analyst served as auditor throughout the research process. As transcripts were completed, the researcher read through each interview once without coding or dividing transcripts

into units so as to acquire an initial sense of each participant's perspective on the WAI. The open coding phase was then undertaken. Interviews were uploaded into the Atlas.ti (MAC OS v.1.01) qualitative data analysis program. The researcher then read through each interview again in detail and divided transcripts into meaning units. Meaning units were identified as segments of text that offered a coherent reflection(s) about a particular concept/phenomenon; these ranged in length from a single clause up to several paragraphs. The number of meaning units within each transcript ranged from 40-151 and totaled 960 across all 12 interviews. Open codes and property statements were generated for each meaning unit throughout this process; sometimes multiple codes or statements were applied to a single unit wherein participants appeared to offer more than one reflection. The goal in this open coding phase was to remain grounded in the data offered by participants, and thus these initial codes were often summarized *in vivo* codes that utilized participants' language directly in the naming of codes. As an example, the first participant offered the following reflection, which was identified as a meaning unit: "I realized that I want to make sure that I'm reading it carefully and making sure that I'm filling it out as accurately as possible" (WAI_08). This unit generated an initial code of 'I want to make sure I'm filling it out as accurately as possible.' The auditor and the researcher initially coded the first interview separately and then, together, compared meaning units and codes generated, and resolved any discrepancies that arose. Possibilities for early higher-order categories were likewise discussed and developed at this point in the auditing process.

After the first interview had been completed in this manner, the codes generated from this transcript were grouped into higher-order categories using the Atlas program.

Relatively quickly into this process, the researcher began to have difficulties visualizing the relationships on the computer, and thus took to printing out initial codes, cutting up these print-outs, organizing codes into categories using these cut-outs, and then returning to Atlas to organize and develop the category structure reflected by the cut-outs.

Using the constant comparison method, nascent categories were continuously evaluated against other categories and subsequently-analyzed codes to ensure each was relatively coherent and distinct. The coding of subsequent interviews was approached in the same manner. Throughout the development of this category structure, the auditor provided advice on the coding process and helped to refine and clarify nascent categories.

The mid-level categories generated from this analysis numbered 129. These 129 categories were likewise constantly compared and were clustered into 17 higher-order categories. These 17 categories comprised the subcategories presented in the final model. Connections among these 17 subcategories were developed, which yielded seven main categories; these are also presented in the final model. Saturation of categories became apparent after the ninth interview was analyzed, meaning that no new categories were proposed as a result of analyzing the subsequent interviews. At this late analytic stage, the auditor served to further refine the category structure and relationships therein, while assisting the researcher in developing and elaborating the reflections on this structure for use in the final report.

Phase Two Method

This phase was designed to address research question two, namely, what concerns do clients identify when rating items on the WAI-SR. Item concerns could include, for example, difficulties with understanding a statement or term, inappropriate response

options, or problems in ascertaining an appropriate response; these areas of focus are described in greater detail in the Analysis section for Phase Two.

Two interview protocols were employed during this phase of the present study: think-aloud and verbal probing. Both protocols were utilized in order to explore each item of the WAI-SR in greater detail than addressed in Phase One. There were two unique samples of five participants each who were randomly assigned to one of the two protocols. Clients were initially alternately assigned to each of the interview conditions based on the date of their initial consent (e.g., the first participant was assigned to verbal probing, the second to think-aloud, the third to verbal probing, etc.), but as one participant withdrawal occurred that required a modification of participant assignment, two think-aloud participants completed interviews back-to-back; as a result, two verbal probing participants were then required to complete interviews back-to-back. All participants were unique to this phase, and thus none had participated in Phase One.

Participants. Participants for Phase Two included 10 clients, five in each interview condition. Six participants were recruited from the YUPPC. The remaining four participants were recruited from the YUPC, two of whom were current York students. Interviews for this phase took place from April to August 2014.

All participants completed the WAI-SR in the research interview session, as both interview protocols required active and immediate reflection on the measure. There were no specific inclusion criteria for this phase. The initial intention was to require that all participants had completed the WAI-SR regularly over treatment, which was the case for nine of the ten participants, but due to schedule constraints and the desire to vary the

sample, the final participant was a client who completed the WAI-SR for the first time in interview (verbal probing protocol).

Participant data for this phase can be found in Table 2. Seven women and three men with a mean age of 25.2 (range = 19-33) took part in this phase. Seven participants had received therapy previously; durations of previous therapies ranged from 1-12 months. Mean WAI-SR item ratings were computed for all participants across all sessions; the range of the number of completed WAI-SR measures among this group was 5-18, with one additional participant having never completed it before. The mean WAI score for this sample overall was 4.42.

Table 2
Phase Two Participant Characteristics

Client ID	Age	Gender	Source	# Sessions[a]	Mean WAI	Protocol[b]
WAI_23	21	F	YUPPC	14	4.78	TA
WAI_24	33	M	YUPPC	13	4.03	VP
WAI_25	19	F	YUPPC	12	3.99	TA
WAI_26	28	F	YUPC	17	4.51	VP
WAI_27	33	F	YUPPC	12	4.36	TA
WAI_28	20	M	YUPPC	21	3.92	TA
WAI_29	24	F	YUPPC	24	4.73	VP
WAI_31	25	M	YUPC	33	4.93	TA
WAI_32	24	F	YUPC	16	None	VP
WAI_33	41	F	YUPC	22	4.52	VP

[a] This column indicates the number of sessions each participant completed prior to interview.
[b] TA = Think-aloud; VP = Verbal Probing

Therapists. There were eight therapists seeing the clients interviewed for this phase. One therapist saw three of the participants, while the remaining seven therapists saw one participant each. Seven therapists were doctoral trainees at York University, while the eighth was a PhD-level psychologist. There were six women and two men. The mean age for these therapists was 30.5 years (range = 27-48), and they had a mean of 4.4 years

(range = 2-8) of therapy experience. Four of these therapists saw at least one client in both Phases One and Two.

Interviews. Interviews during this phase lasted 25-90 minutes. All were audio-recorded and transcribed. All interviews were completed within 24 hours of participants' most recent treatment sessions. Seven of the 10 interviews took place following participants' final sessions while the remaining three participants were continuing in treatment past the dates of interview. The procedures of each interview method are distinct and are presented below.

Interview method one: Think-aloud protocol. The first interview protocol was the think-aloud procedure during which clients were asked to narrate their internal processes when completing each item on the WAI-SR. Other than the initial invitation to narrate, the interviewer did not ask specific probes or solicit specific feedback from participants during their active reflections. He did occasionally redirect participants back to the task with prompts as needed (e.g., "Tell me what you're thinking") (Priede & Farrall, 2011). In order to orient participants to the task of thinking aloud, they were asked to undertake two exercises prior to interview. They were encouraged to elaborate on their responses during this orientation, as necessary. The first orienting exercise was as follows:

> Try to visualize the place where you live and think about how many windows there are there. As you count the windows, tell me what you're seeing, feeling and thinking (adapted from Willis, 2005, p. 44; as cited in Priede & Farrall, 2011, p. 276).

Participants were then asked the second introductory exercise that helped to familiarize them with thinking aloud while considering WAI-style questions. This second exercise was

the following: "On a scale of one to five, five being 'Always' and one being 'Seldom,' how frequently do you agree with the following statement: My therapist and I understand each other." This sample statement was taken from the original, long-form WAI (Horvath & Greenberg, 1989), but it is not included on the WAI-SR and had thus never been encountered by participants. As with the interview proper, there were no interviewer probes during this second introductory exercise, other than to remind or encourage the participants to think aloud while considering the item and their response to it. After participants' uninterrupted narratives, probes (spontaneous and emergent probes (Beatty & Willis, 2007)) were employed at the interviewer's discretion to further elucidate participants' experiences with each item.

Interview method two: Verbal probing protocol. These interviews utilized a structured protocol that also asked participants to go through each item on the WAI-SR in interview. Following Karabenick and colleagues (2007), participants were encouraged to read every item aloud and then were asked three core questions for each item on the measure: 1) What is this question trying to find out from you?; 2) Which answer did you choose as the right answer for you?; and 3) Why did you choose that answer? Participants were asked follow-up probes (e.g., "Can you tell me more about that?" or "Can you give me an example?") if the initial question(s) failed to yield a sufficiently detailed or clear answer.

Analysis. Interviews were transcribed and analyzed chronologically. Transcripts were content coded as to whether clients indicated difficulties with any of the items on the measure and, if so, which type of problems were identified. Transcripts produced using both interview protocols were subjected to the same analytic method. The pre-defined coding framework, the Classification Coding Scheme (CCS), that was utilized in this phase

was described in Rothgeb and colleagues (2005) and corresponded to the four-stage Tourangeau-Rasinski cognitive model of survey responses.

The CCS framework is broken down hierarchically into three tiers of codes. The highest-level codes correspond to the four stages in the Tourangeau-Rasinski model (Question Comprehension, Information Retrieval, Judgment, and Response Selection). The CCS hierarchy is presented in Table 3 below. The higher-order categories of 'Comprehension and Communication' and 'Response Selection' have two tiers of code levels subsumed within them, while 'Retrieval from Memory' and 'Judgment and Evaluation' have only a single tier each. It was the lowest level of codes in each category that was applied to interview transcripts. Lower-level codes were the primary unit of analysis for this phase; however, higher-order groupings were used for descriptive purposes.

Once all interviews were completed and transcribed, the researcher uploaded them into the Atlas.ti software. First, the researcher read through each interview to familiarize himself with participants' reports. Then, he read through each transcript chronologically to identify passages in which participants described a potential problem with an item. When a potential problem was identified, the researcher delineated the relevant passage in Atlas.ti (akin to segmenting meaning units in the GT analyses of Phase One) and assigned a relevant first-level CCS code (e.g., 'Undefined term(s)/vague term') to the excerpt. All interviews were coded in this fashion. Unlike Phase One analyses wherein all interview data were coded, only those interview portions in which participants clearly identified a potential problem were formally assigned a code in this analysis.

Each CCS code thus had one or more interview excerpts with which it was associated. The researcher then grouped the excerpts associated with each code according

to the WAI-SR item(s) on which they reflected. All excerpts associated with one code and one WAI-SR item were then reviewed to identify the specific problem that had been

Table 3
Classification Coding Scheme (CCS) Structure

Comprehension and Communication

- *Interviewer Difficulties*
  - Inaccurate instructions
  - Complicated instructions
  - Difficult for interviewer to administer
- *Question Content*
  - Vague topic/unclear question
  - Complex topic
  - Topic carried over from earlier question
  - Undefined term(s)/vague term
- *Question Structure*
  - Transition needed
  - Unclear respondent instruction
  - Question too long
  - Complex or awkward syntax
  - Erroneous assumption
  - Several questions
- *Reference*
  - Reference period carried over from earlier question
  - Undefined reference period
  - Unanchored or rolling reference period

Retrieval from Memory

- Shortage of memory cues
- High detail required or information unavailable
- Long recall period or long reference period

Judgment and Evaluation

- Complex estimation, difficult mental arithmetic required
- Potentially sensitive or desirability bias

Response Selection

- *Response Terminology*
  - Undefined term(s)
  - Vague term(s)
- *Response Units*
  - Responses use wrong or mismatched units
  - Unclear to respondent what the response options are
- *Response Structure*
  - Overlapping response categories
  - Missing response categories

reported by participants. For example, when the 'Undefined term(s)/vague term' code was applied to excerpts related to Item 7 ("I feel that _____ (my therapist) appreciates me"), those excerpts were reviewed to identify the specific term that participants had described as being undefined or vague (the word 'appreciates' in this case). All problems were investigated in this manner, and the findings were collated according to WAI-SR item number.

A second coder was then included to undertake reliability coding of the problem codes; this coder was an advanced doctoral student in Psychology at York University. The researcher met with the reliability coder initially to provide an orientation to the CCS structure and to practice coding examples together. The researcher then provided the second coder with interview excerpts from the first three interviews, which represented roughly one-third of the excerpts in the total sample. The second coder applied relevant CCS codes to each of the excerpts, with the researcher being available for general coding consultation throughout the process. Coded excerpts were returned to the researcher and a reliability estimate across both coders was derived.

Phase Three Method

Phase Three consisted of the synthesis and elaboration of the findings from Phases One and Two in the service of developing a more comprehensive understanding of the WAI-SR from clients' perspectives. Lincoln and Guba (2000) noted the importance of triangulation as a way of enhancing the validity of findings by demonstrating the complementarity and convergence of results gleaned from multiple methods. To this end, findings in the present study were triangulated based on an approach described in Meijer et al. (2002). Following Miles and Huberman (1994), Meijer and colleagues (2002)

described triangulation as a process by which researchers intended to develop further support for a finding by demonstrating that complementary investigations agreed with initial results.

The present study employed methodological triangulation by including multiple interview procedures and analytic methods and utilized an intuitive triangulation procedure for integrating findings across methods (Meijer et al., 2002). Following Smaling (1987; as cited in Meijer et al., 2002), Meijer and colleagues described intuitive triangulation as a means by which a researcher builds connections among results gleaned from multiple data sources without utilizing a formal procedure. This intuitive process was applied in the present study to ascertain the degree to which findings converged or diverged from and elaborated on each other across phases. Farmer and colleagues (2006) offered another example of a triangulation procedure at work, exemplifying the ways in which researchers could reflect on and represent points of convergence, divergence, and silence across different phases in a study.

In an effort to enhance the triangulation of findings and allow clients' perspectives to speak for themselves, a brief single case analysis of a Phase Two participant's think-aloud interview is also included. This interview affords deeper access into one participant's spontaneous, real-time rating processes and serves to enhance the representation of clients' approaches to completing the WAI-SR. In light of this, 'triangulation' undertaken in the present work intentionally begins to approach, though does not wholly embrace and implement, what Ellingson (2009, 2014) refers to as 'crystallization.' Ellingson (2009) offered the crystal, rather than the triangle, as a more apt metaphor for qualitative research. She posited that the multi-faceted, three-dimensional crystal was more reflective

of postmodern qualitative work than the simpler, two dimensional triangle. As Denzin (2012) noted, Ellingson,

> Endorses a postmodern form, asserting that the central image for qualitative inquiry is the crystal, multiple lenses, not the triangle. She sees crystallization as embodying an energizing, unruly discourse, drawing raw energy from artful science and scientific artwork. (p. 83)

Ellingson (2014) describes crystallized texts as incorporating several integral elements: 1) They generate knowledge using a deep, thick description and interpretation of findings; 2) They use a range of methods of analysis from different points along the qualitative spectrum (e.g., art-based, constructivist, positivist/postpositivist); 3) They include multiple writing genres (e.g., conventional academic prose, poetry, song); 4) They incorporate significant reflexivity on the part of the researcher; and 5) They deny positivist/modernist claims to singular, objective truth. Developing a crystallized text allows a researcher to reflect on, interact with, and convey the complex interactions among diverse narratives and research methods utilized in qualitative inquiry.

In the service of pointing in the direction of a crystallized text, Phases One and Two were offered as two distinct, albeit related, lenses through which clients' perspectives of the WAI were acquired, interpreted, and represented. Phase Three sought to engage with and integrate these findings by offering a third lens through which to view and represent clients' perspectives. This lens included the summary and integration of findings developed throughout Phases One and Two, as well as unfiltered access to one participant's perspective by reproducing and analyzing the think-aloud narrative component of her interview. While not intended to be a fully realized crystallized text in the vein of Ellingson

(2014), Phase Three is meant to complement and enhance the two lenses represented throughout Phases One and Two.

Research Results

Phase One Results

Phase One focused on addressing the first research question, namely, "How do clients understand the WAI and on what do they base their ratings?" To address this question, semi-structured interviews were undertaken with clients in ongoing psychotherapy. Grounded theory method was employed to analyze participants' interview narratives. The resulting category structure is presented below.

Qualitative Analyses. ***Grounded theory code structure.*** The code structure developed over the course of the Grounded Theory analyses completed in this phase can be seen in Figure 1. The core category, 'Rolling on the right track,' is presented first. Main categories were divided according to two main headings: *Translating* and *Relating Together*; both main headings reflect complementary elements identified in the core category, as described below. Main categories are presented next with subcategories subsumed under each.

Core category. The core category **Rolling on the Right Track** reflects two themes identified throughout clients' accounts about engaging with the WAI-SR. On the one hand, participants stated that they utilized the measure as a means with which to convey to therapists and/or their clinical supervisors how therapy and the work therein was progressing (i.e., as a measure of perceived treatment effectiveness). They used their responses as a way to present the degree to which the therapeutic work was 'on the right track' and progressing in the desired direction towards achieving treatment goals. One participant summarized this fundamental perspective by conveying her sense that the WAI

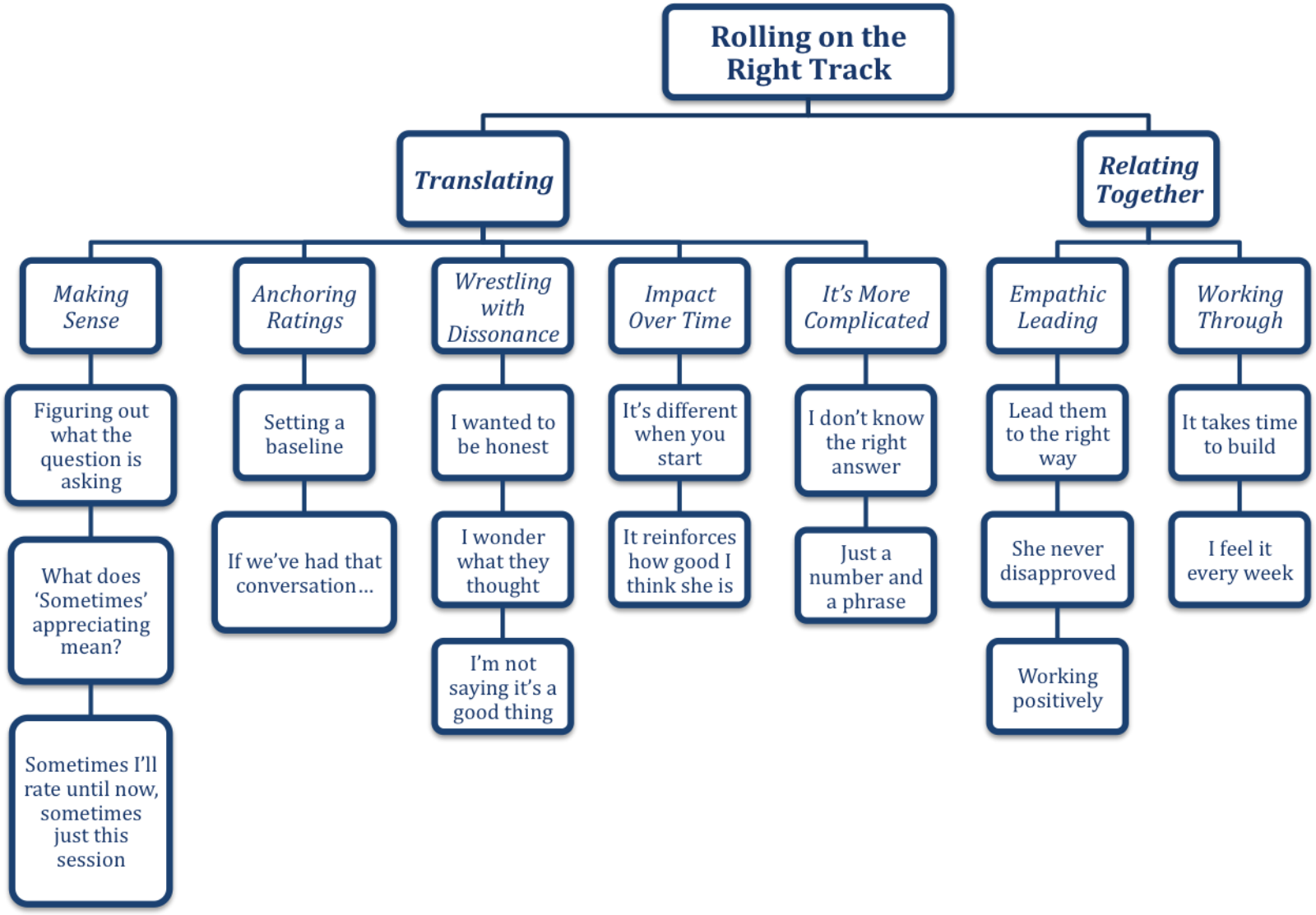

Figure 1. *Phase One Category Structure*

was a measure intended, "to see if we're on the right track on the things that I want" (WAI_15).

The majority of participants in this sample identified their therapies and therapists as positive and helpful, and many participants wanted to ensure that they conveyed these positive feelings to therapists. For example, one participant described this positive alliance in the following terms: "I circled 'Always' because I feel like I derive a lot of positive benefit both emotionally and intellectually from talking to [my therapist]" (WAI_20). As such, participants viewed WAI ratings as an important and active, albeit not always ideal, means with which to deliver treatment feedback. In short, participants seemed to view the measure as a means with which to gauge, and more focally convey, that therapy was working well (or was not working well, as was the case for several participants). In the majority of cases in which participants identified positive treatment work, early indications of relational support from therapists and early treatment gains served to increase the degree to which participants believed treatment to be effective and meaningful (i.e., these early experiences increased the sense of 'rolling' along an appropriate and helpful track). These facets of **Rolling on the Right Track** are described throughout the main categories subsumed under the *Translating* heading.

The core category **Rolling on the Right Track** also highlights participants' reflections of the salient elements of the therapeutic work and the connections with therapists that helped them to make progress (i.e., factors that aided them in getting and remaining 'on the right track'). Participants elucidated beliefs in two focal elements their therapists offered to them: an ability to guide participants with helpful suggestions and the use of a non-judgmental stance and other relationship factors. In many cases, specific

participant interview accounts included factors that were explicitly queried by the WAI, though there were notable exceptions to this in which participants described variables that were either not queried by the WAI or were queried on the WAI but which participants deemed to be unimportant. Thus individual participants' perspectives as to what being 'on the right track' entailed were variable. Depictions of these helpful alliance factors are captured within main categories under the *Relating Together* heading.

Translating and Relating Together categories. The seven main categories and their respective subcategories were subsumed under two headings: *Translating* and *Relating Together*. The former heading comprises categories that primarily highlight participants' engagement with the WAI and the processes they used to evaluate, rate, and communicate feedback about their experiences in treatment. These categories describe the active processes through which participants worked to understand, personalize, and respond to the measure. As the primary emphasis during these interviews was on the impact of the rating process itself, rather than on relationship or treatment effectiveness per se, *Translating* categories predominate in these analyses.

The *Relating Together* categories describe participant narratives focused on productive treatment and relational factors. These categories offer insight into participants' reflections on important elements in treatment, including change processes and helpful elements in their connections with therapists. The most salient factor in participants' positive relationships was the therapist's non-judgmental style.

Translating. These categories emphasize participants' engagement with the measure itself, and how they translated experiences in sessions to interpret and respond to WAI items. This process required that participants actively translate their experiences in

treatment into data that could be represented numerically on the scale. This process of engagement with the measure was often a dynamic and active one, requiring that participants interpret the WAI's items and distill their experiences in therapy into appropriate ratings. These processes involved developing an understanding of the measure and its elements (*Making Sense*) and devising an awareness of how to identify relevant ratings (*Anchoring Ratings*). The rating processes also included grappling with the meaning and impact of those ratings on therapists or supervisors and treatment (*Wrestling With Dissonance*), the uncertainty of meaning and use of the ratings (*Impact Over Time*), and the ultimate purpose and focus of the measure (*It's More Complicated*). Participants' interview excerpts utilized within each main and subcategory will be presented in the sections below that introduce each of the categories.

Making Sense. This main category reflects how participants interpreted and made sense of the elements of the WAI. In working to make sense of this measure, participants engaged with the WAI at three levels: the items, the rating scale, and the reference period queried by the measure. A number of participants described a process of active engagement with the measure as a whole in which they worked to understand the terms in certain items, such as the term 'likes' in item 3 ("I believe _____ likes me"), or the applicability of an item to their particular therapies (e.g., the therapist disapproving of something the participant had done, as queried by item 9). Most participants found the majority of items on the scale to be reasonable and intelligible. Participants also engaged with the anchors on the scale, working to understand the different levels. They did this by personalizing interpretations of the qualitative scale descriptions, such as 'Sometimes, or by interpreting the scale as dichotomous rather than graded when applied to certain items.

Finally, participants worked to understand the reference period queried by the measure (e.g., whether their ratings were based only on the most recent session or treatment as a whole), noting it to be ambiguous; some participants highlighted tendencies to vary their approaches to this ambiguity, rating variously based on only the most recent session or some combination of previous ones. All participants contributed to this main category. It is comprised of three subcategories: 1) *Figuring out what the question is asking;* 2) *What does 'Sometimes' appreciating someone mean?*; and 3) *Sometimes I'll rate until now, sometimes just this session.*

Figuring out what the question is asking. This subcategory reflects participants' understandings of and engagements with the WAI-SR items; all participants contributed reflections to this subcategory. Participants described their impressions indicating that the majority of the items were clear and intelligible. For example, one participant shared the following: "I get the point of all of them, like why they would be asked, so it seems pretty straight-forward" (WAI_22); however, nine other participants identified elements in at least one item that they reported as being difficult to understand in some respect. While the majority of participants described at least one difficulty with one or more items, the particular items they identified as challenging varied across participants.

Several participants reported challenges with understanding and responding to item 3 ("I believe _____ likes me"), describing the term 'likes' in this item as "ambiguous" (WAI_19) and noting they may not "have a clear definition of what 'like' is" (WAI_08) in the context of psychotherapy. Other participants noted that although rating whether their therapist liked them felt "awkward" (WAI_21), "weird at first" (WAI_12) or less important

than other therapist qualities (WAI_16, WAI_19), they did not have difficulties understanding the term 'likes' as used in the item.

Others identified challenges with some of the questions that referred to setting goals in early sessions, and whether they were applicable at such an early point in therapy.[2] Several participants described the first several sessions as being about letting therapists get to understand them as people, rather than about explicitly setting goals for the therapy, which made answering such items difficult in the early stages. One participant described this in the following way: "Only by session four you might think of a possible solution and something to change, so in the beginning question one's not very applicable" (WAI_11). Another described it in this way:

> In the beginning, I was mainly talking and saying everything and telling my whole life story and then later, like midway or after a couple of the beginning sessions, then we started collaborating [on goals]. (WAI_15)

Here is another perspective: "We didn't have any goals to begin with, so I didn't know how to rate that on a scale" (WAI_08).

Some participants noted difficulties with item 9 (i.e., "I feel that _____ cares about me even when I do things he/she does not approve of"), as they could not recall any situations in which they had done something that would have elicited their therapists' disapproval; thus they had difficulties identifying a salient memory or feeling to which they should refer in answering this question. In the absence of a clear situational reference,

---

[2] When citing participants' reflections about goals-focused items, I am not indicating that participants were referring specifically to items included on the Goals subscale. The majority of participants conflated Goals and Tasks questions; thus reflections about items about goals did not necessarily line up with the actual subscales on the measure. This tendency is described in further detail elsewhere in this section.

these participants noted that they imagined how their therapists might have reacted had they done something that prompted the therapists' disapproval. In all cases in this sample, participants reported believing that therapists would not have been judgmental or disapproving of participants' decisions; however, this inferred lack of negative judgment did not always result in participants rating 'Always' for this item.

Terms in two additional items were identified as confusing by two participants: one participant identified the term 'new way' in item 2 and another identified the word 'appreciates' in item 7 as "ambiguous." There were no other questions identified by this study sample as being problematic with regards to phrasing or item referents; thus participants found the majority of WAI items to be clear and understandable.

What does 'Sometimes' appreciating someone mean? This subcategory captures how participants engaged with the levels of the scale itself (e.g., 'Always,' 'Very Often,' etc.); eight participants contributed to this subcategory. A number of participants reported on their tendencies to modify or customize their interpretations of the scale; this process occurred in different ways. Some noted that they would at times re-interpret the verbal descriptors on the points of the scale. For example, one participant said the rating of 'Fairly Often' actually meant 'Sometimes' to her (WAI_11). Several participants also noted that they read 'Seldom,' the lowest rating on the scale, as actually reading 'Never,' which is exemplified in the following: "In terms of 'Seldom,' I really look at that as being really almost like a 'Never'" (WAI_16). Some participants identified that their active engagement with the scale led them to reinterpret some of the items as only including 'Yes' or 'No' response options, namely, replying with only 'Always' or 'Seldom,' since the graded scale

made little sense to them on some items. For example, one participant described this problem in response to items 3 and 7:

> Both are yes or no questions, like do I feel [my therapist] likes me or no [my therapist] doesn't like me, or does she appreciate me or does she not, it's not like oh for the first half of session she liked me but last week she didn't, it's more like yes she does or no she doesn't. (WAI_19)

Two participants indicated that they informally equated the scale levels with percentages of confidence in the items' statements (i.e., 'Always' equated to 100% certainty in the item, 'Very Often' equated to 75-80% certainty, and so on). These participants adapted this scale to be more applicable to their interpretation of certain items. For example, WAI_11 summarized several of these modifications in the following way:

> So a 1 I wouldn't give because 1 means we never do it…2is also pretty low, and 3 means 'Sometimes' for me, even though 2 is 'Sometimes,' and then 4 is usually like I give it on a 85-90%, and 5 would be 100%, which would be like total confidence. (WAI_11)

One additional participant noted that she occasionally modified the numerical scale by, for example, providing 4.5 as a rating. She noted that she seldom used this option but did so when she believed that the existing five-point scale was limited in its ability to convey her beliefs.

In addition to their tendency to adapt the scale levels, four participants reflected their belief that shifts of one or two points along the five-point scale did not represent significant changes. For these participants, whether they chose a 3 or a 4 (or a 4 or a 5, etc.) on a particular item was of limited significance. The two following excerpts exemplify this

perspective: "It's not really that important if I do a three or a four, it's not a huge difference for me" (WAI_21). Another participant conveyed a similar perspective in the following:

> I follow my first instinct [after rating 4/Very Often], even if after the fact I'm like maybe it should have been 'Always', really there's not much of a difference between the two, so I'm not coming up short anywhere. (WAI_18)

Sometimes I'll rate until now, sometimes just this session. Participants described different interpretations of the scope of treatment that should be encapsulated by their ratings. Eight participants offered narratives described in this subcategory. Many participants noted that they primarily used the most recently completed session as the referent for WAI ratings, whereas several others identified that they typically based ratings on their sense of the general relationship and work over the course of treatment.

For some participants, the choice of the approach would depend on the particular item. This hybrid approach meant that they might rate based on the recent session if relevant, but in the event that a recent session was not as positive or was less relevant with regards to an item's statement, they would rate based on their overall experiences in treatment. In short, the participants noted that if an item's referent was present in the most recent session, they would use that to answer the item, but if such a referent had not occurred in the recent session, they would answer based on their overall experience. A participant described this approach in the following excerpt:

> It depends on what happened in the session, sometimes I'll do until now and sometimes I'll do just that session, especially when something sticks out like if we're moving towards mutually agreed upon goals, like if we have a certain goal that day

and I agreed to it then okay, but then if we didn't discuss goals much, then I'm just like, let's see what happened this whole time. (WAI_15)

Another identified the same approach in the following manner: "I feel like I was more inclined to score her higher even if I wasn't really certain with that session because she had helped me previously" (WAI_19). Different participants addressed the implicit ambiguity of the period covered by the measure by developing and utilizing various approaches. It is noteworthy that no participants expressed explicit confusion about the intended reference period queried by the measure.

Anchoring Ratings. This main category of responses encapsulates participants' perspectives on how they represent the quality of their therapeutic relationships through interaction with the WAI. Eleven participants contributed reflections. Participants' narratives conveyed elements of the ways in which they decided on ratings in general. They described tendencies to identify default start positions, or baselines, on the scale. For example, some participants identified their tendency to begin by assuming something queried had *always* been true in treatment (i.e., set their baseline rating to 5/Always) and sought exceptions that might dispute their assumption and thus might decrease their ratings. In the process of providing ratings, participants would then work to anchor an appropriate rating using memories from specific events and/or feelings from sessions, such as, for example, an explicit memory of a time a participant's therapist had said, "I appreciate you," thus providing an anchor to respond to item 7 ("I feel that _____ appreciates me). This category is comprised of two subcategories: 1) *Setting a baseline;* and 2) *If we've had that conversation…*

Setting a baseline. This subcategory reflects some participants' descriptions of anchoring their ratings to a baseline start position on the scale and then gauging their responses from this point; five participants offered descriptions included in this subcategory. Participants utilized these baseline positions to help provide a starting point for ratings from which they would judge the appropriate response. For participants who noted that they engaged with the scale in this way, several identified the 'Fairly Often' as their baseline, while others identified the 'Always' as their anchor.

Participants who reported that the 'Fairly Often' level was their beginning point highlighted their beliefs that this point on the scale represented a relatively neutral and safe point from which to begin (i.e., neither too positive nor too negative). One participant described it in the following way: "I just lean towards the middle, a safe bet…it's like you can't really get much on it, it's neutral, so had I put something lower or higher, it would give off something and I didn't want that" (WAI_15). This participant began with the 'Fairly Often' rating and noted her sense that, at times, she wanted to convey neutrality in her ratings in doing so. Another participant reflected on his approach: "If I don't know [what to rate] or I'm unsure, I usually put three" (WAI_14).

Several other participants identified the 'Always' rating as being their baseline for some of the items. They reported that they would begin by assuming that 'Always' was the most accurate choice and search their memories for evidence that disproved their inference that the item's statement had *always* been true. For example, WAI_19 described this approach:

> Because she never disapproved or showed disapproval or showed disappointment or anything of the sort, then I felt that she always liked me and always appreciated

> me because if she didn't, then wouldn't she express some form of disapproval or uncertainty when I said certain things? (WAI_19)

Reflections from other participants further exemplified this process: "I don't know if my therapist does [respect me], like I assume, 'cause I didn't give them a reason to not respect me or appreciate me" (WAI_14). Another noted it in the following way: "As far as the respect each other and stuff, maybe I put 'Always' just because we didn't get into a fight the first session, so that went well I guess" (WAI_12). And another:

> When it gets into thinking about how we deal with my goals, it gets a little complicated because I have to think back to times when I didn't feel, like for [item] 12 especially, there were times when I didn't feel like we were dealing with it as quickly as I would have preferred, so that's why I put 'Very Often.' (WAI_18)

These participants would in essence begin by assuming something had *always* been true and would then attempt to search for explicit memories that would serve as exceptions to their assumption. In the absence of those exceptions, they would provide 'Always' as the rating of choice; if exceptions were discovered, the number of those exceptions could dictate which lower rating was more accurate (e.g., a single exception might decrease the rating to 'Very Often,' whereas several might decrease it to 'Fairly Often').

If we've had that conversation… This subcategory highlights differences in approaches to the types of information participants used to anchor their ratings; nine participants provided reflections included herein. Participants described two general approaches with the majority of contributors reporting that they used some combination of the two. One tendency used explicit memories based on in-session experiences to ground ratings. This approach was focused on how particular interactions would impact individual

item ratings, the exact content of which varied by question. These references to explicit memories provided greater ease with which ratings could be completed and heightened confidence in the accuracy of such ratings. For example, WAI_13 described her approach to answering item 9 ("I feel that _____ cares about me even when I do things he/she does not approve of") in the following way: "I think about times when I felt worried about sharing something and then I was trying to think back, did I feel like she was judging me in any way when I said something."

In contrast, the second approach was noted to involve tendencies to provide ratings based on an intuitive or experiential sense of how the session or sessions had gone. Clients endorsing this approach tended to reflect on how they felt about their therapists and their progress and then offered ratings that they believed represented this intuitive sense. WAI_22 described this approach in the following way, citing that she generally provided 5/Always ratings on the measure as a whole based on, "the satisfaction that you get at the end of the session and how you feel when you look back at it. If it seems like everything went pretty well, and you were satisfied, then why not give the best."

Most clients contributing to this category endorsed the use of both strategies with respect to different questions, although they expressed differing degrees with which they did so. In the following excerpt, WAI_08 exemplified both strategies by contrasting her ability to rate one item about which she and her therapist had spoken with another one about which she had not had an explicit conversation:

> She's said a couple of times that she appreciates working with me and appreciates the opportunity she has to help me so because we've had that conversation, it [item

#7] makes sense to me, but we've never had the conversation about whether we like each other or not, so I'm not clear on that one [item #3]. (WAI_08)[3]

WAI_08 would later go on to note that she responded to item 3 based on her therapist's "positive energy" towards her and her sense that questions about 'appreciation' and 'liking' were similar, thus should be given similarly high ratings. In short, to help answer an item that she found difficult to understand, she utilized both her experiential sense of her therapist's feelings towards her and an explicit memory used to rate a similarly phrased item.

Wrestling With Dissonance*.* This category reflects the degree to which participants experienced comfort or tension in providing feedback on the WAI, and all participants contributed to this category. Two participants described feeling significant tension in 'wrestling' with the degree of honesty in their ratings and intentionally inflated ratings, while the remainder identified the importance of answering honestly and accurately, even if such an approach resulted in some degree of discomfort. Some participants further identified beliefs that the WAI queried and implicitly prioritized therapy elements that were not integral for therapy success, such as the emphasis on collaborative goal setting or having the therapist like his/her client. This category includes three subcategories: 1) *I wanted to be honest*; 2) *I wonder what they thought;* and 3) *I'm not saying it's a good thing.*

I wanted to be honest. Ten participants contributed narratives reflected in this section. Eight of these ten participants expressed their desire to convey accurate ratings to therapists and clinical teams and reported that they were honest when providing their WAI

[3] These item numbers as presented are reflective of their numbering on the WAI-SR. This participant was the only one who completed the older WAI-S.

ratings. They conveyed the hope that ratings would not lead to confusion or misinterpretation and wanted to ensure their feedback was as accurate and honest as possible. Participants noted the hope that their mood or other extraneous variables did not impact the ratings they provided.

> I'm used to coming in with the mindset of just trying to answer these things honestly, not trying to manipulate things based on what I think somebody wants to hear, just try to base it on what I think is my honest experience. (WAI_20)

Two other participants noted that they inflated their ratings occasionally due to concerns about the negative effect that low ratings might have on therapists' confidence, course grade, or on the effectiveness of treatment. For example:

> Last time I put a four down [for #1] because we didn't really talk about changes, but I still put a four down because I think overall I guess, I rated it a four maybe not specific to that exact session, because I didn't want to put a one because I perceive a one to be poor. (WAI_11)

This participant continued: "She's great, but the way I did it is more towards making her seem great, which she is" (WAI_11). Another participant shared a related perspective:

> I always gave her positive ratings [for #2], but there were times where I kind of went back afterwards and was like, oh I don't know if I got anything from that, but I never reflected that in my answer (WAI_13)

Participant WAI_13 in particular frequently referenced her tendency to provide glowing ratings that might not have been completely accurate. She later reflected on her experience of 'wrestling' with several possible rating choices:

> Those were where I really wrestled 'cause I felt 'Very Often' is more realistic but then I would feel like it's also negative because in our sessions I *always* did feel that this person liked me, but I just don't think that's a realistic goal either, to have somebody always like you. (WAI_13)

This participant described a strongly positive relationship while also noting reluctance to provide anything but the highest ratings.

I wonder what they thought. Seven participants from the sample provided relevant reflections for this subcategory that comprised accounts regarding participants' comfort or discomfort with completing ratings. Several of these participants noted that providing ratings on the WAI resulted in no discomfort, while others reported that they felt mild discomfort with completing the WAI. This discomfort stemmed from concerns about their therapists and/or therapists' supervisors reading the completed measures and potentially having negative reactions or misinterpreting ratings, as described in greater detail above. For example, WAI_12 noted tension in saying that his therapist 'Always' liked him after having had only one session:

> I know she's probably gonna think it's weird for me to say "I believe [my therapist] always likes me" if she's only met me for one session, so if I put 'Always' she's gonna say 'oh he's just saying that' so I don't know what to put. (WAI_12)

This client noted that his discomfort with this question dissipated after the first one or two sessions.

Another participant noted the following concerns about being perceived as "mean" in her ratings:

> If I put that the therapist isn't working, we're not agreeing or something, we're not collaborating on goals or I'm not agreeing with them, it just feels a bit mean to say even if it was true. This person's trying to help me. (WAI_15)

One participant in particular expressed a high degree of discomfort with providing ratings. She conveyed that part of this discomfort was due to being asked to speak on her therapist's behalf by, for example, conveying whether she believed her therapist respected her:

> It's hard for me to answer this question (#5) because I don't know if the other person feels respected 'cause I'm doing a lot of the talking, so I try to be consciously respectful and…during my sessions I felt like I was being respected so I'm saying yes. I can't say 'Always' because I don't know if they feel respected. (WAI_21)

This participant's completion of the WAI resulted in recurring feelings of mild guilt. She described her belief in the importance of being honest in her ratings and noted feeling guilty as a result of providing any ratings below 'Always.' She described ambiguity in the use of the ratings, specifically whether her therapist was being graded on the basis of WAI ratings, and her resultant guilt in the following excerpt:

> I feel guilty when I'm giving [my therapist] 3s and 4s because I don't know what that means. I don't know if this questionnaire means absolutely nothing or if she's being graded on it…it would help ease the guilt of my answers knowing that I don't have an impact on somebody else's grades. (WAI_21)

I'm not saying it's a good thing. This subcategory represents nine participants' perspectives on the degree to which the WAI-SR content reflected their own beliefs about important factors in therapy. A number of participants highlighted their perspectives about

key elements in therapeutic work as in line with the WAI-SR messaging (e.g., mutual respect, agreement on goals and targets, helping produce change, and so on). For example, WAI_19 relayed the general importance of her therapist's guidance, lack of negative evaluation, and positive support:

> "[My therapist] was always respectful, she's very aware of the changes I want to make, maybe changes I didn't see and maybe enlightened me to things I wanted, but she never disapproved if I said I'm not comfortable with that. I never felt like she was pushy, she was always suggestive, and a positive support."

Additional, specific areas of agreement are described further in the *Empathic Leading* category below.

Four participants described at least one discrepancy in their beliefs about important elements in therapy and those being queried on the WAI-SR. The participants reflected on their sense that some of the items asked questions about areas that, for these participants, were peripheral, such as items that emphasized collaborative goal setting. They noted that they might provide high ratings on certain items yet hold personal beliefs that the elements being queried were not relevant for treatment. For example, one participant shared her reflections about questions focused on goal setting:

> I understood it was important to have mutually agreed upon goals, and I did see that during therapy, she wanted to come up with those goals for me, with me, but even though I said yes 'Very Often' or 'Always', it didn't necessarily mean that I actually liked that. I know it was important, but I'm not saying that it's a good thing when I say 'Very Often' or 'Always,' I'm just saying that's what it is. (WAI_16)

Two other participants echoed this disagreement about the implicit importance of 'collaboration' referenced throughout the WAI-SR. These participants noted that they wanted therapists who took the role of expert, eschewing collaboration in the service of allowing therapists the freedom to lead. One participant described this further in the following excerpt:

> The question she tends to ask a lot is, do you want to do this, is this where you want to take it this time, and I don't know if it's because the supervisor's there because I know she has a set, what she'd like to do, and I'm ok with that too…you're the professional, if you think that's ok I'd like to go along with it 'cause you obviously know more than me about the topic. (WAI_11)

In contrast to this participant, others described their appreciation of their therapists' emphasis on mutual agreement on topics covered and techniques utilized in treatment.

Participant WAI_16 also identified discordance regarding being liked by her therapist. She reported having a strong belief that her therapist liked her, thus providing a high rating on item 3 yet noted her sense that her therapist's liking of her was irrelevant for treatment progress; what mattered for this participant was whether her therapist was technically proficient in facilitating change, which included maintaining mutual respect but did not extend to 'liking' one another.

Three participants had particular concerns about questions that referenced 'changes' in treatment, and they noted how such explicit changes were not central for them. One participant described this in the following manner:

> A lot of [the questions] are about how I'm able to change and how we looked at ways of changing, and I'm just aware for rating these that they kind of assume that in

> every session I'll change in a big way, whereas…it's not like I have an epiphany every time. (WAI_14)

Another reflected on 'changes' in this way:

> The idea of changes, which is also related to goals, maybe it's more applicable to more people in the study than to me, but it isn't as relevant to me at least in terms of accomplishing changes. I mean yes there's always things that we want to accomplish in whatever situation we are in, there's always ways to look at that, but sometimes it's more about how to go about that in the best way possible without necessarily a focus on only being goal-driven. (WAI_20)

These participants tended to interpret the word 'changes' as denoting significant life changes, rather than more day-to-day changes. It was this interpretation against which they appeared to be reacting and with which they expressed disagreement.

Impact Over Time. This category reflects all 12 participants' reflections regarding their connections with the measure over time. Included in the subcategory *It's different when you start*, participants offered accounts relating a changing degree of engagement with the measure, often describing an initial period necessitating greater interpretation and processing of the measure followed by a period of consistency and predictability in their responses; some participants, however, conveyed that their relationship and approach to completing the WAI had not changed notably over multiple completions. This main category also incorporates participant reflections as to the degree to which those multiple completions may or may not have had a salient impact on treatment and/or participants' relationships with their therapists. These accounts were incorporated within the subcategory *It reinforces how good I think she is*.

It's different when you start. All participants reflected on the increased familiarity and predictability of their ratings over multiple sessions and multiple completions of the WAI. Four of these participants noted that their views of and engagement with the WAI remained consistent over successive completions. For example, one expressed this consistency in the following way: "I don't think it's something I really have to sit down and think about for a long time, like it's quick and I can think about it easily…I've never had any problems, I've never been stuck" (WAI_22).

The remaining eight participants reported that their understandings of and approaches to the measure had changed to some degree over treatment. This latter group of participants described their encounters with the measure in the first few sessions as resulting in some confusion as to what certain items were referencing or how they were "supposed" to rate. For example: "I didn't understand what these questions were asking of me initially" (WAI_08). Most participants reported that this increased familiarity with the measure resulted in an increase in the speed with which they completed the WAI after the first few sessions. Participants explained this as being the result of developing a stable understanding of the questions they would be asked, as well as, for many, an increased predictability of the ratings they would provide. One participant noted the following:

> After the second or third or fourth time my decision was kind of made about how I was going to answer these, and because it's the same evaluation and she hasn't changed and she's still a great therapist, then my answers are going to be the same. (WAI_11)

And another participant reflected on this theme in the following way:

> I filled them out a lot more quickly [over time]…'cause I knew what to expect, it's not that I was just filling it out without thinking, but I knew I felt comfortable around her, so this is what I'm going to fill out. (WAI_18)

This process of familiarizing oneself with the WAI involved elements of personalizing the scale as well. Participants reported that they worked to personalize their approach to the questionnaire when items seemed confusing or inapplicable to them at first. Once this initial work was completed after the first few sessions, participants stated that they were able to respond more quickly to the items. WAI_20 explained this process in the following excerpt:

> I took more time to think about the questions the first time and then come up with, well for example, some of the questions which contain phrases that are not as applicable to me personally, then you just think about what's a better way for me to answer this so that I can still fulfill the purpose of the questionnaire and what it's trying to get at. (WAI_20)

In addition to actively adjusting item phrasing to make it more personally relevant, as the participant above described, others noted tendencies to modify their interpretation of the rating scale itself (e.g., converting the frequency descriptors to percentages of agreement with each statement) or to adjust how they read the verbal descriptors (e.g., interpreting 'Sometimes' to mean 'Never'), such as when WAI_11 described, "1 [Seldom] means that we never do it…3 [Fairly Often] means Sometimes to me." Some of these strategies were described more thoroughly in the *Making Sense* main category above as well.

It reinforces how good I think she is. Nine participants contributed to this subcategory. Most of these participants identified their belief that completing the WAI had

no impact on their relationship or work in treatment with their therapists; they viewed the completion of the measure as a necessary yet personally inconsequential element of their courses of therapy.

Two participants expressed worries about the impact that low ratings might have on their therapists and/or the treatment progress. These participants noted concerns that therapists might have negative reactions if they saw clients providing low ratings (note that the definition of 'low' varied among participants). Participants suggested that these low ratings might result in therapists feeling upset and/or have deleterious effects on treatment progress. Participants who expressed these concerns also highlighted tendencies to inflate their ratings on occasion, and some of them conveyed having genuinely positive relationships with their therapists yet still perceived the need to inflate ratings (e.g., wanted to rate an item as 'Very Often' but perceived pressure to rate 'Always'). One participant articulated her worries about a negative reaction in the following excerpt:

> I could see it having a very negative effect on future sessions if I rated that I don't think they like me or I don't think they respect me, so a lot of the ones where you put in the person's name, I guess I almost felt obligated to give them a four or five. (WAI_13)

Three participants expressed their beliefs that completing the WAI resulted in a positive impact on their views about and progress in treatment. These participants noted that completing the WAI served to reinforce pre-existing positive feelings they held towards their therapists. For example, one noted that giving her therapist high ratings,

> Reaffirms that she's good after every session, when you fill something out and instead of giving negative feedback, because there's no question here that I would give negative feedback, it just reinforces how good I think she is. (WAI_11)

Another noted the positive impact that her ratings might have on her therapist's confidence: "If she sees [my ratings], then she knows how much I do like working with her, and it shows her how effective she's being" (WAI_18).

It's More Complicated. This category of participant reflections focused on the degree to which participants believed the WAI accurately and appropriately captured their experiences in their relationships with their therapists. Participants reported on how well the measure identified salient facets of their relationships in therapy, and reflected their understandings of the purpose of the measure, as well as their beliefs that it was an evaluative tool to gauge the effectiveness of their therapists. This category includes reflections from all participants, and is comprised of two subcategories: 1) *Just a number and a phrase;* and 2) *I don't know the right answer.*

Just a number and a phrase. Seven participants offered relevant reflections for this subcategory. These participants reflected on their sense of whether the WAI captured their experiences in the relationship, as reflected in this subcategory. Many participants voiced their beliefs that, in general, they perceived the WAI to be a fine measure of their therapeutic work. Some participants stated views that the measure was reductionist and too generic to capture the nuances of their particular relationships and treatment work with their therapists. They likewise noted their sense that "real" feedback needed to be given to their therapists (or therapists' supervisors) verbally. One participant communicated this perspective in the following reflection:

> It's kind of weird 'cause…you can't rate a friendship or rate a relationship or rate your therapist based on numbers, it's like they're, I don't know, it just doesn't seem right. (WAI_19)

This participant went on to describe her preference for direct verbal communication of any feedback with her therapist in session, which seemed more "right" to her than conveying feedback indirectly through any kind of questionnaires. Another participant reflected on his belief that the measure was reductionist:

> Like 'I'm clearer now as to how I'll be able to change' like that's a really long process and I can't really answer it with just 'Yes' or 'Very Often,' it's just more complicated than this, it's hard to describe using numbers. (WAI_14)

Five of these participants identified that, while a quantitative evaluation of a relationship may not allow for perfect representation, it was a generally efficient and reasonable method with which to assess a complex phenomenon. The following participant conveyed it as follows: It,

> Is a tricky thing to do to come up with a quantitative measurement of someone's personal and possibly emotional experience, right? So I mean yeah, I am more or less used to it and I don't think it's necessarily ideal, but it's difficult to come up with any way to have a better method. (WAI_20)

One participant noted the perceived limits of the measure and her belief that the measure would need to assess both parties in the relationship to be more accurate.[4]

[4] Recall that therapists were not completing a clinician version of the WAI in this sample, and thus only WAI ratings from clients were obtained.

> It would capture the relationship, but at the same time it's just a standard question for this, like my thought process for this is it's just a form to capture my thoughts about how I'm feeling about the relationship, but it would be one-sided. If you're to talk to my therapist, it would be different, I mean not different, but you can't just capture a relationship from one side. (WAI_12)

These participants expressed perspectives that relationships were complicated and multifaceted, and described being aware of the relative limitations of attempting to measure such relationships with numbers; however, most explicitly conveyed understanding that it remained an efficient means with which to do so.

I don't know the right answer. Nine participants contributed to the identification of this subcategory. Four participants described being unsure how their WAI ratings were used and by whom. One noted her belief that supervisors utilized ratings to gauge participants' confidence in therapists' abilities. Two others reported their belief that only therapists used the ratings. Another noted thinking that ratings were used solely for research purposes. One of these participants highlighted this ambiguity in the following excerpt:

> In the beginning, you don't know who's looking over these, so you don't know whether you want to put a number one or a number five, which is 'Seldom' or 'Always,' so you don't know…what it means if you put a one or a five, like is it gonna be reflecting for the therapist, who's overall a great therapist, so you don't want it to reflect poorly on them if you put a one. (WAI_11)

This participant further identified some of the ambiguity in the rating of certain items and highlighted difficulties in identifying what she 'should' rate. She conveyed this sentiment in the following:

> She's great, but then, well the way I did it is more towards making her seem great, which she is, but a lot of these questions are very abstract, and I don't really know the right answer to it. (WAI_11)

Another participant expressed concerns with ambiguity as to how to rate items that ask her to adopt her therapist's perspective. She articulated the following in regards to her ability to make judgments of her therapist's perspectives:

> Blank appreciates me (#7), and he and I respect each other (#5), and I believe that blank likes me (#3) 'cause some of them like I have no idea. Sometimes I don't know what they're thinking so I would just view myself, and if I was the therapist what I would think of myself, which obviously in the beginning wasn't that positive, but then I was kind of getting reassurance from him that he wasn't judging me or anything, then those were a little bit easier to answer. (WAI_15)

This participant highlighted that the confusion in the process of rating, and in adopting her therapist's perspective in particular, became less taxing over time.

Despite some degree of confusion regarding the specific ways in which WAI ratings were utilized, participants understood the measure to be an evaluation of how their therapists were doing in therapy and whether they were progressing demonstrably in treatment. Participants described the measure as a general means with which to evaluate how successful and helpful the treatment had been, as well as how helpful their therapist had been throughout the process. One participant compared the WAI-SR to evaluations of

university professors found on the website 'Rate my Professor' as a means with which to measure progress. Another expressed a related sentiment in the following way: "It's about the sessions that I have with [my therapist], and that this is important to determining how well we're doing together" (WAI_16).

Three participants conveyed the view that the WAI was comprised of essentially two clusters of questions. One component identified was focused on participants' comfort level in session with their therapists, while the other focused on how productive and helpful sessions had been. One participant reflected on this perspective as follows:

> Half of [the items] is [*sic*] how she's making me feel, if I'm comfortable in the sessions, and then [the other half] would be do I think she's doing a good job, 'am I liking the way she's taking it in the process we're going through.' (WAI_11)

Another participant described it in this way:

> There could be like two questions on here. One could be, '[my therapist] and I respect each other' and [the second could be] 'Are we working together productively to give new ways of looking at my problem and helping me accomplish those changes?'…Those seem to be the same questions just reworded a bunch of times for the entire 12 questions. (WAI_14)

No participants in this sample offered reflections indicating that they perceived the measure to be comprised of three subscales (Bond, Goals, Tasks), and as such they tended to combine items on the Goals and Tasks subscales.

Relating Together. These main categories and subcategories reflected narratives emphasizing the centrality of elements at work in participants' relationships with their therapists and the treatment process. These categories encompass salient facets of

participants' therapeutic relationships that were key for them, as described in the *Empathic Leading* category. Participants further reflected on their own processes in treatment and how making treatment gains impacted these relationships; these reflections are integrated in the *Working Through* category.

Empathic Leading. This main category encompassed descriptions from all 12 participants of elements of positive working relationships with their therapists. Participants described these positive relationships as involving therapists acting as guides throughout the treatment process and helping to assist participants in navigating problems, such as when WAI_15 reflected that it is "important for the therapist not to just listen, it's better to ask questions to guide in the right way." A number of participants identified what they viewed as important ingredients in their positive working relationships with their therapists that include therapists' active leading during the treatment process and refraining from passing negative judgment on client experiences. Participant WAI_19 described these latter experiences in the following reflection: "Whether I was angry or I cried or I smiled, it didn't matter how I reacted, she always supported that." This category includes three subcategories: 1) *Lead them to the right way;* 2) *She never disapproved;* and 3) *Working positively.*

Lead them to the right way. Eight participants spoke about the importance of therapists acting as guides throughout the treatment process. In part, this meant recognizing and deferring to therapists' expertise and technical proficiency with mental health treatments and participants' desires to follow therapists' leads. The therapist-as-expert was of primary importance in these reflections. For example:

> After I listened to her talk and reflect back to me, I saw she had techniques and so for me, that was more important than the liking and the caring and the hand holding. I really didn't need that. (WAI_16).

Participants' descriptions in this vein connected with concerns about an over-emphasis on 'collaboration' running throughout the WAI-SR, as described in the *Wrestling With Dissonance* main category above.

Having therapists lead the way meant they could help to make flexible suggestions and facilitate participants' identification of goals. Therapists' roles were described as helping to develop clients' awareness and refinement of treatment targets and facilitate clients' acquisition of relevant skills. One described it here: "Another good thing is letting the patient or client come up with things on their own, and if they're not going that way, lead them to the right way" (WAI_15). Another reflected on her therapist's guidance in the following excerpt:

> She's very aware of changes I want to make, maybe changes I didn't see, but suggested and maybe enlightened me to things that I wanted but [she] never disapproved if I was like, oh I'm not comfortable with that. (WAI_19)

This client went on to describe the mechanism through which her therapist helped to guide her in sessions: "She would kind of weigh my options for me, or help me to weigh my options myself, like give me those tools to do it myself" (WAI_19).

She never disapproved. A consistent relational element identified by eight participants was reflections about therapists' non-judgmental support. These participants reported that their therapists helped develop comfort and trust in sessions by avoiding judgmental statements or behaviours. Participants suggested that this non-judgmental

orientation allowed them the freedom to be open and vulnerable about their feelings or behaviours without concern about receiving disapproval from therapists. One participant reflected on the impact of her therapist's non-judgmental stance in the following excerpt:

> Sometimes I can come off sounding a little crazy, but she doesn't look at me as though I'm crazy or her responses don't indicate that she feels that what I'm saying is a little eccentric, so that helps me feel more comfortable again. (WAI_08)

Another participant shared a similar reflection:

> My therapist made me feel that the way I'm handling myself in certain situations and relationships that I had, that I wasn't at fault 'cause I used to think that. Basically he made me kind of feel like I wasn't a bad person. (WAI_15)

While emphasizing the centrality of non-judgment as a relational tool, both of the above excerpts also help to highlight the impact that therapists' non-judgmental support can have in reducing self-criticism.

Working positively. All 12 participants contributed to this category. The majority of participants described having positive relationships with their therapists in general which they reported as being reflected in their ratings. These participants highlighted their beliefs that their therapists liked, respected, and cared about them. Participants emphasized the importance of both comfort with their therapists and progress achieved via treatment as key elements in working positively. These factors led to higher ratings in general:

> I feel like if you've built that connection, you're almost always gonna rate them higher because they were helpful, because if I walked into my first therapy session and I just didn't feel a click or something like that, then maybe I would have answered this differently, but because it was there, I felt that comfort and like

comfortable to open up a little more, then maybe I was more inclined to give her higher scores (WAI_19)

Comfort experienced in sessions and the helpfulness of those sessions were central for this participant, as for others.

In addition to describing the relational elements, five participants also described the importance of more technical aspects demonstrated by their therapists, such as therapists' abilities to listen, summarize, and collaboratively develop goals, or therapists' preparedness for sessions. For one participant, having her therapist summarize or reflect back to her what she had said was an integral element of their work. She described this in the following exchange in which she is reflecting on positive qualities exemplified by her therapist:

P: [She should know] how to drag things out of you, so artful at doing that and [my therapist] was, sometimes I listen to what I said and she was able to turn that back around and make it look like it's actually coming from me, and then like, oh lord, did I say that?

I: What was it like when she did that?

P: I felt challenged, and I like someone to challenge me, that was important to me, not just think what I say because I am a bullshitter sometimes, and I have a certain way of doing things, and it's the right way, and obviously it hasn't worked for me. (WAI_16)

This was a participant for whom the idea of her therapist 'liking' her seemed irrelevant and secondary to more change-focused strategies, hence the centrality of the more change-focused strategies described in her reflection.

Three participants reported that another element of working positively included therapist friendliness. Two participants noted the importance of this facet of the relationship while also highlighting the need for the relationship to remain professional: "It's still a client-patient relationship. Even if I talked to her as though she's my friend because I feel like I can be open and honest with her, at the same time I shouldn't cross that line" (WAI_08). Another client reflected on the importance of the friendly-yet-professional relationship in answering item 3 and her discomfort with being overly positive:

> I don't want to…creep her out. Like by saying yes [i.e., rating a 5], so I'm trying to maintain that professional relationship and respect her boundaries as well as myself…I feel like she does care, but I think 'Fairly Often' is safe because if I say like she 'Very Often' cares, it's kind of, I don't want to cross that client therapist boundary, so I'm trying to keep it neutral (WAI_21)

Working Through. This main category reflected 10 participants' sense of increasing trust and comfort with their therapists and the treatment process, which was strongly connected with experiencing personal change over successive weeks in treatment. The majority of clients reflected on this trend during interviews, and these reflections suggested two subcategories: 1) *It takes time to build;* and 2) *I feel it every week.*

It takes time to build. Six participants reflected on their increasing sense of trust and confidence in their therapists and the treatment process as treatment progressed. There were no participants who conveyed that they were as trusting of therapists and the process at the beginning as they were later in treatment. One participant described part of her process of increasing trust as involving her experience in working on therapeutic targets with her therapist:

> Because I didn't know her so well and I was feeling apprehensive about [treatment], I didn't feel that immediate trust, but as we worked through some issues together, then I started opening up more, and that's when I started maybe grading or writing a number that was higher on the scale because I was able to trust her a lot more than I was in session one. (WAI_08)[5]

Another participant identified increasing confidence in her therapist and treatment in the following way:

> It's very hard to fill out this questionnaire, more so at the beginning because, not that there's not a connection, but you don't know yourself how you feel about therapy, so it's like asking, can you do the course evaluation at the beginning of class? (WAI_19)

A third participant described her sense that she was apprehensive about all aspects of treatment in the early stage, which changed over her time in therapy:

> In the beginning I was kind of skeptical if anything would work. Say we had a goal, later on [I thought] it was a good thing, but in the beginning [I thought] this is not going to work, so I was doubtful of everything. (WAI_15)

One participant highlighted the fundamental importance that her trust in her therapist had on her WAI ratings in the following way: Trust "impacts all the numbers because if I didn't trust her I'd feel very reserved and I think all the numbers would change" (WAI_11).

I feel it every week. Eight participants reflected on how they perceived changes in their lives as a result of their work with therapists in treatment. They highlighted a number

---

[5] This participant was the only one who had completed the initial short-form version of the WAI (Tracy & Kokatovic, 1989). This original version has a question that explicitly queries clients' trust in their therapist; this question was omitted in the WAI-SR.

of mechanisms through which these changes had been undertaken (e.g., reduction in anxiety due to exposure work, changes in thinking styles, increases in self-esteem). One participant outlined her sense of relief experienced over her course of treatment in the following way:

> Coming in to this, my mom used to make fun of me 'cause I used to be like Hercules, the mountain on my shoulders, but now I feel that's been lifted and just that feeling alone, I completely feel at peace. This scoring perfectly reflects how I'm feeling, I feel physically, mentally, emotionally a kind of relief, so that's all because of the sessions, because maybe the relationship that the therapist and I have developed over the course of sessions, and I feel like if I hadn't trusted her or I hadn't opened up to her as much as I am now, I don't think we would have made that improvement at all. (WAI_08)

Another participant described how her self-concept had improved over the course of treatment and how her initially low self-esteem was tied to assumptions that her therapist likewise saw her negatively; this assumption about her therapist's perspectives about her also led to assumptions that her therapist did not approve of things she had been doing. The following excerpt conveyed this participant's initial beliefs and how these changed over treatment:

> In the beginning it was mostly me talking, and I guess it wasn't until the first time I ever got reassurance, I wasn't putting high numbers, especially the one about this person cares about me even though I do things they don't approve of (#9), because I personally don't approve of them so why would somebody else? So there was a bit of bias there…because I used to think negatively about myself. (WAI_15)

What this participant refers to as 'reassurance' from her therapist is presented as an integral element in her treatment progress. This participant also conveyed that answering question nine described above was challenging initially because she would put herself in her therapist's shoes, assuming that he was judging her negatively for behaviours about which she herself felt negatively, and thus rated the item based on her own sense of herself.

Some participants communicated that their ratings on the items relating to 'changes' and therapeutic 'goals' were based on their perceptions of their personal changes over the course of treatment. In essence, these participants noted that because they had made meaningful changes over the course of therapy, they must by definition have agreed on relevant goals in treatment. For example, one participant articulated this connection in the following way:

> Because it's close to the end, and I've actually seen the changes, [my ratings] are more on the positive side. For example, the first question itself, it's asking whether she and I agree on what I'll need to do to improve my situation. For me when I was reading that question, clearly we've made an improvement, so obviously we agree or we wouldn't have reached this. (WAI_08)

Another participant reflected her belief that her answers to questions about therapy goals seemed obvious to her because she had begun to experience changes in therapy. This participant noted that her referents for such questions shifted over therapy, as her initial ratings were related to her expectations that therapy would be helpful, whereas her later ratings were grounded in changes she had actually experienced: Questions about goals seemed obvious,

> Because I felt that was the point of therapy for me, to help me make changes to achieve things. Closer to the end or middle of therapy, [answers to goals-focused questions] were still kind of self-evident because [my therapist] was helping me to achieve those goals, so it was like, at the beginning this is the point of therapy and at mid/end of therapy it was like [my therapist] *is* doing this because I feel it every week. (WAI_19)

This participant later went on to elaborate on her approach to completing questions about goals in earlier sessions before she had experienced changes: Early goals-focused ratings were based on,

> General things we'd spoken about and potential changes that could be made to achieve those goals and, like I said, that was the point of therapy so I assumed [my therapist] would do that. (WAI_19)

Another participant reflected on her experience that her time in therapy yielded increases in her hope and optimism about treatment progress, which in turn yielded higher ratings on items connected to therapy goals and tasks. She described it in the following manner:

> Things started working, and I could kind of see why my therapist was saying some things…so I guess I was more optimistic. I had more hope, so I feel like the questions on goals and what is important to work on, I agreed with them. Most of the things were pretty much the same, but questions like that I was more likely to give fours and fives. (WAI_15)

Summary. Interviews conducted with 12 participants in Phase One utilized a semi-structured interview protocol and were analyzed using Grounded Theory methods (Corbin

& Strauss, 2008). Participants' accounts acquired during Phase One resulted in a number of relevant themes that were identified in seven main categories that together led to the development of the core category, **Rolling on the Right Track**. This core category encapsulated two complementary views expressed by participants. The first regarded the function and purpose of the WAI-SR, namely, that it was perceived to be and was utilized as a means with which to evaluate and communicate how effective courses of treatment and participants' therapists had been (i.e., a way to communicate whether participants and therapists were "rolling" towards changes and "on the right track" together in treatment). This core category is also grounded in participants' descriptions of those therapeutic elements that allowed them to be "on the right track" in their respective treatments.

This core category was comprised of seven main categories that elucidated ideas contained therein. These main categories, divided across two broad dimensions of ***Translating*** and ***Relating Together***, were as follows: 1) *Making Sense*; 2) *Anchoring Ratings*; 3) *Wrestling With Dissonance*; 4) *Impact Over Time*; 5) *It's More Complicated*; 6) *Empathic Leading*; and 7) *Working Through*. Each of these main categories will be summarized briefly below.

Translating. *Making Sense*. This main category captured participants' reflections highlighting their investments in understanding the facets of the WAI when providing ratings. Participant accounts reflected on this process of 'making sense' of the measure and its elements, describing an active engagement in working to understand occasionally confusing components in the WAI. Accounts centred on interpreting three specific components of the WAI: the phrasing of the items (captured in the subcategory *'Figuring out what the question is asking'*), the points

on the rating scale ('*What does 'Sometimes' appreciating someone mean?*'), and the reference period queried by the measure as a whole and/or specific items ('*Sometimes I'll rate until now, sometimes just this session'*). Some participants provided specific examples of confusing items or terms, such as with the term 'likes' in item 3 ("I believe _____ likes me"). Some items were also described as being inapplicable for participants in general. Participants also highlighted questions about the rating scale itself, occasionally misunderstanding or personalizing the qualitative descriptors or viewing the scale as dichotomous rather than graded. Finally, participants engaged with the reportedly ambiguous period queried on each item. Some responded to the items by referring only to the recently completed session when providing ratings, others reflected on elements of previous sessions when rating, and a number incorporated elements of both strategies when responding.

Anchoring Ratings. This main category highlighted the ways in which participants anchored their ratings, both to a baseline position on the rating scale and to specific memories and/or feelings identified in session(s). Participants identified baseline ratings from which they gauged their ratings, such as by assuming something had *always* been true in therapy and then attempted to identify exceptions to this inference (*Setting a baseline*). Participants also sought to anchor their ratings in relevant data by referencing explicit memories of in-session encounters and/or feelings regarding their therapists in order to judge which ratings were appropriate (*If we've had that conversation…*).

Wrestling With Dissonance. This category included narratives that highlighted the degrees to which participants experienced tension while rating. Participants offered accounts about how accurate and honest they were when rating therapists and courses of treatment (*I wanted to be honest*) with most participants identifying their beliefs in the importance of being honest when answering items. For some, such honesty led to feelings of discomfort, while others stated that they felt quite comfortable completing the WAI (*I wonder what they thought*). Tension was further evident in some participant accounts that noted inferences that the WAI might, at times, prioritize elements in treatment that were not of importance for everyone, such as setting goals collaboratively with one's therapist. Most participants noted, though, that their perspectives about important therapeutic factors were highlighted in the WAI (*I'm not saying it's a good thing*).

Impact Over Time. This main category encompassed participant accounts focused on how their engagement with the WAI changed or influenced them and their treatment over time. A key finding here pertained to participants' awareness that some had initial confusion with elements of the measure, but that their investment in making sense of the measure and subsequent regular completions of it led to a more quickly accessible, predictable and stable pattern of responses (*It's different when you start*). A second subcategory incorporated participant reflections as to whether they believed the completion of the measure had changed their engagement with their therapists and/or courses of treatment. The majority of participants did not believe their ratings had an effect; however, some believed ratings may have impacted treatment, either negatively by resulting in therapists'

hurt feelings at seeing 'low' ratings or positively by serving to reinforce participants' already positive feelings towards therapists and treatment (*It reinforces how good I think she is*).

It's More Complicated. This category was comprised of participant reflections about how they did or did not understand the purpose of their WAI ratings and the ways in which they were utilized throughout treatment (*I don't know the right answer)*. Numerous participants voiced beliefs that the WAI was a measure of general therapeutic effectiveness while noting a lack of clarity about how and by whom the results were used. Participants also offered reflections on how the WAI was viewed as a generally relevant, efficient means with which to gather information about clients' views of therapy, albeit one that could not fully convey the nuances of a multifaceted relationship (*Just a number and a phrase*).

Relating Together. *Empathic Leading.* This main category included participant accounts that conveyed what they considered to be key elements in a positive working relationship. Participants highlighted the importance of therapists' approaches in helping to guide participants over their courses of treatment by working to set reasonable goals, utilizing helpful techniques, and offering suggestions about how to make change (*Lead them to the right way*). Participants also identified a key relationship element as being a non-judgmental stance adopted by therapists that allowed participants to feel more comfortable and helped to develop trust (*She never disapproved*). Finally, participants offered accounts on the strength of their connections with therapists and treatment. The majority identified

having a strong, positive connection and finding treatment to be helpful in making changes (*Working positively*).

Working Through. This final main category encapsulated participant reflections that highlighted how their relationships and work with therapists changed over time. Most participants indicated that their sense of trust and confidence in their therapists developed together over successive sessions (*It takes time to build)*. A majority of participants reflected on their beliefs that treatment had been helpful for them in reaching their goals. Some of these participants also highlighted the connection between experiencing changes in treatment and subsequent goals-focused ratings on the WAI and noted that their goals-related ratings were based on experiencing change while in treatment (*I feel it every week*).

Phase Two Results

Ten participants were interviewed in Phase Two to gather item-specific reflections regarding the intelligibility of each item and participants' methods from which they derived and communicated their ratings. Two interview protocols were utilized in this phase: a semi-structured verbal probing protocol (n = 5) and an unstructured think-aloud procedure (n = 5). Participant accounts were analyzed using the Classification Coding Scheme (Rothgeb et al., 2005).

A brief quantitative description of the codes applied to Phase Two interviews is presented first. This summary offers an overview of the frequency with which all codes were applied to the entirety of the sample. The overview is followed by an item-specific summary that outlines the frequency with which each code family was applied to each of the WAI-SR items, as well as the frequency with which individual participants identified

concerns with each of the WAI-SR items. This quantitative summary is not intended to be exhaustive. The more detailed and elaborated qualitative summary follows and is intended to provide a more thorough presentation of the findings.

Quantitative Summary. ***Reliability coding.*** After the primary coder had finished coding all transcripts in this sample, the first three of the 10 coded transcripts were provided to a secondary coder to assess reliability of coding. These three transcripts represented nearly one-third of the total code sample (29.8%; 28/94 codes). While there are 27 possible CCS codes, only seven were utilized by the primary and secondary coders in this reliability sample. After determining the code agreement between the coders, a Cohen's unweighted kappa of 0.77 was computed using statistical software (VassarStats). A kappa of 0.77 is considered to be of moderate strength, which is adequate for interpretive purposes (McHugh, 2012).

Overview. A total of 111 problem codes were applied across the sample; a summary can be found in Table D4 in the Appendix. The majority of codes (94) were applied to item-specific reflections, and, in some cases, a single code was applied to a single reflection that referenced multiple items simultaneously. Seventeen codes were applied to sections in which participants were making overall reflections on the measure, and these were, therefore, not specific to particular items. Table D4 summarizes the frequency of each problem code as applied to an item-specific utterance across the full sample. For some interviews, multiple codes were applied to the same identified problem (e.g., one participant stated her opinion that she found the word 'appreciates' in item 7 to be vague at

several points during her interview, and the 'Undefined/vague term' code was applied and counted in each instance).[6]

'Comprehension and Communication' difficulties accounted for 71.3% of codes (67/94). Of these Comprehension and Communication codes, 'Question Content' concerns were the most frequently noted, comprising 74.6% (50/67) of the sample, followed by 'Question Structure' concerns at 19.4% (13/67) and 'Reference Period' at 6.0% (4/67). No interviewer or instruction difficulties were noted in this sample.

The next most frequently applied code group was 'Judgment and Evaluation' concerns, which comprised 14.9% (14/94) of the code total. This group consisted entirely of 'Potentially Sensitive/Desirability Bias' concerns.

Finally, 'Response Selection' difficulties constituted 13.8% (13/94) of the total codes. This code group was comprised of 'Response Structure' codes at 61.5% (8/13), 'Response Units' codes at 30.8% (4/13), and 'Response Terminology' codes at 7.7% (1/13). The code group, 'Retrieval from Memory,' was not evidenced in this study.

***Item-specific summary*.** Codes applied to specific item numbers were tabulated, and the frequency with which codes were applied to specific items can be found in Table D5 in the Appendix. Each item on the WAI-SR had an average of nine codes applied to it (range = 0-25). Similarly, the number of coded reflections contributed by each participant to specific WAI-SR items was calculated and is presented in Table D6 in the Appendix. Table D6 includes only codes applied to participant reflections that were specifically

[6] It is for this reason that the total number of codes (94) described in this general overview is lower than the total (108) described in the following item-specific summary section.

targeted towards one (or more) WAI-SR item. Participants contributed an average of 10.8 specific item codes during each interview (range = 0-28).

The majority of item-specific concerns were raised in regards to items comprising the Bond subscale (57.4%; 62/108); this subscale includes items 3, 5, 7, and 9. Codes applied to the Tasks scale encompassed 24.1% (26/108) of the total; this scale includes items 1, 2, 10, and 12. The Goals scale accounted for the remaining 18.5% (20/108) of the codes and is comprised of items 4, 6, 8, and 11.

In terms of codes applied to specific items, item 7 ("I feel that _____ appreciates me") had the most codes applied to it, accounting for 23.1% (25/108) of the item-specific total. The majority of the codes applied to item 7 pertained to 'Question Content' concerns (88.0%; 22/25).

Item12 ("I believe the way in which we are working on my problem is correct") contributed the next highest proportion, accounting for a further 17.6% (19/108) of item codes; difficulties with 'Question Content' again made up the majority of these codes (57.9%; 11/19).

Item 3 was the next most frequently coded item, contributing an additional 16.7% (18/108) of codes. In the case of item 3, the majority of concerns again pertained to 'Question Content' (66.7%; 12/18).

Item 9 had 13.9% (15/108) of codes applied to it, more than half of which referenced concerns with 'Question Structure' (53.3%; 8/15). The remaining WAI-SR items had fewer than 10 codes applied to them. Table D5 presents the codes applied to these remaining items.

Not included in this breakdown were the 17 codes that were applied to participants' utterances that reflected only on the global measure, rather than referencing a specific item. Nine of these global codes were categorized as instances of 'Potentially Sensitive/Desirability Bias.' A further five identified concerns with the 'Reference Period,' two identified issues with the 'Response Units,' and one noted 'Response Terminology' difficulties. These global concerns are not a focus for Phase Two item-specific results, and thus are not discussed further.

Qualitative Results. This summary lists and describes all participants' concerns as regards the specific items on the WAI-SR. It describes these concerns listed by each item and grouped according to the subscales. Excerpts from interviews are provided throughout to illustrate participants' reflections.

***Bond subscale*.** Codes applied to items on the Bond scale represented the majority of codes in this sample (57.4%; 62/108). Nine participants contributed reflections to items on the Bond subscale.

Item 7 – "I feel that _____ (my therapist) appreciates me." This item had 25 codes applied, accounting for 23.1% (25/108) of the total codes; this item thus had the most codes applied to it of any item. Seven participants provided remarks that contributed reflections on this item (see Appendix).

Participant comments centred on several key concerns. The majority of responses pertained to the word 'appreciates' and characterized the term itself as undefined or vague, as well as characterizing the question as unclear in general due to this term; six participants identified this concern. Participants highlighted their difficulties in understanding the term as utilized in the therapeutic context and noted questions as to

why or how their therapist would appreciate them. They identified confusion with the term 'appreciates' in light of believing it to be more appropriate in a personal relationship rather than a professional one. One participant highlighted this confusion as follows: "Why would he appreciate me because he's providing something for me and I'm not necessarily providing something for him" (WAI_23). Another participant grappled more extensively with this term and the criteria with which she could respond to this item:

> I don't know what the criteria would be for 'appreciates me', like what would make him not appreciate me, would it be me saying something that he didn't approve of? I've not been shown any indication, well I did miss a couple of sessions, so does he not appreciate me 'cause I didn't come to those sessions, I cancelled them in advance, but like I don't know what the criteria for 'appreciates' is. (WAI_27)

Several participants noted that the confusion with this term led them to replace the term with a synonym that allowed them to comprehend and answer the question, such as the word 'respects' used in item 3. These participants thus answered item 7 in the same manner as item 3.

Two participants highlighted their tendency to rate this item highly because they were unclear as to the question's meaning and did not want a low rating to reflect poorly on their therapists. Thus, a desirability bias became a concern as a direct result of perceived ambiguity in the question. One participant described it in the following exchange:

> I: Because the question doesn't quite make sense, how do you choose which rating to give?
>
> P: I give a five because it sounds good. (WAI_23)

Another noted her tendency to give a five for this question because "I don't want it to look bad on [my therapist]" (WAI_29).

One participant raised her belief that the item covered a complex topic that was difficult to assess accurately. She related her difficulties in answering this question with her belief that it was a challenging item to rate due to difficulties in identifying her therapist's genuine feelings, as opposed to her perceiving her therapist as "just doing her job." She explained this concern in the following excerpt:

> It was easy to understand, a bit more difficult to answer because some questions ask how she feels or how I perceive it, which, because it's a professional, it's difficult to differentiate what are someone's genuine feelings or just how they've learned to do their job. (WAI_32)

A final participant noted his belief that this item was a question requiring a yes or no answer, not a graded one; he thus believed that the response units (the 1-5 Likert scale) were inappropriate. This participant did not identify the rating scale as inappropriate across the entirety of the measure.

Item 3 – "I believe _____ (my therapist) likes me." This item was the next most frequently coded one, adding 18 codes and accounting for 16.7% (18/108) of the total code pool. Five participants contributed to codes on this item (see Appendix).

Two participants highlighted difficulties in understanding the term 'likes' and thought it reflective of a complex topic. These participants identified that gauging the degree to which their therapists liked them was a challenging proposition. They noted being unclear as to how they could go about effectively assessing this statement.

Two participants reflected on the possibility of desirability bias affecting ratings on this item. These participants described concerns that the item was somewhat leading and pulled for high ratings. One participant described how he would be reluctant to indicate explicitly if he believed his therapist did not like him, but he might express that implicitly by discontinuing therapy:

> I: Can you say more about the tension, or if you did feel like she didn't like you, how you would respond to the question?
>
> P: I wouldn't say that, I just wouldn't show up the next time. (WAI_24)

This participant would convey his belief by terminating early, but he would not be likely to express this in his rating on this item.

One participant raised a concern about the rating scale units. He identified confusion about how a therapist might vary the frequency of their 'liking' for clients; he described it as more of a yes or no question, as he also noted about item 7: "They don't really mean anything 'cause you see someone every week, so what does that mean, like somebody likes you three weeks as opposed to four weeks in a row? It just didn't make much sense to me" (WAI_24).

Item 9 – "I feel that _____ (my therapist) cares about me even when I do things he/she does not approve of." This was the next most frequently coded item in the Bond subscale with 15 codes identified, accounting for 13.9% (15/108) of codes. Seven participants provided relevant reflections on this item (see Appendix).

Five participants described an erroneous assumption on the part of this item, namely, that there existed a situation of relevance in which participants had disclosed something to their therapists about which he or she would have disapproved. These

participants identified that no such situation had taken place to their knowledge, and thus the question was not applicable to them. One participant described this problem in the following way: "See, this one I don't even feel that it applies to me in my situation 'cause I don't think I do anything, so that's why I always just pick 'Always' because I don't know" (WAI_27). One of these participants reported that this item had been inapplicable for her for the majority of treatment but had become relevant during a recent session in which she disclosed something about which she was worried her therapist would not approve. In short, the implicit assumption in the item had become accurate and relevant for her by the time of interview, but it had been erroneous for the preceding several months.

One of the above participants stated that her preference would have been to avoid answering this item, if possible, considering it did not apply to her; she described several other items in the same way. She also identified confusion as to the reference period covered by the measure, being unclear as to whether the measure referred to the entirety of treatment or was aimed at specific sessions. She described it as follows: "If this questionnaire is for specific sessions, then those particular ones on certain sessions if I had the choice not to answer them, I wouldn't, but as a whole that's why I put 'Fairly Often" (WAI_27); she thus noted that she defaulted to providing a rating of 3/Fairly Often because she was not presented with a formal option to mark an item as inapplicable. Another participant offered the same perspective, citing his preference to answer the item as 'not applicable.' As a result, this participant never actually rated item 9 on any of his completed measures and wrote 'N/A' beside it each time; he completed every other item consistently.

Item 5 – "_____ (my therapist) and I respect each other." This is the final item in the Bond subscale, and accounted for only 3.7% (4/108) of all codes; these codes came from two participants (see Appendix).

As identified above regarding item 3, one participant described his belief that the response options for this item were not appropriate as a graded rating scale, but rather reflected a dichotomous state of his therapist either respecting him or not respecting him. He thus suggested that response units would more appropriately be 'Yes' or 'No.' This same participant indicated his confusion about the purpose of the question and its relevance for his therapy.

Another participant indicated her recent tendency to write in the margins beside this item to explain her rating. She reflected her belief about the importance of adding such notes to explain her rating of 4/Very Often as being due to her own lateness to session and not disrespect on the part of her therapist. She articulated this in the following exchange (as she was in the process of completing the measure with the interviewer):

P: I was late again (laughs), so I'd give it a four, and I put a little note saying it's my fault

I: What is it about the note writing that you feel is important?

P: To show that ownership, that it's not [my therapist's] fault, it's mine

I: And so if you were to put the four down now, how would that feel? I guess if you were to try and see, does it fit?

P: Makes it feel uncomfortable (laughs) just because I was late, and I'm putting it out there, but it also makes me wary because I don't want the surveyor, or whoever collects this, to think that [my therapist] is disrespectful in any way. (WAI_29)[7]

This exchange highlighted the participant's belief that her ability to explain her ratings was insufficient, hence her desire to elaborate on the reasoning behind her rating to ensure her therapist was not viewed poorly. The participant also described her approach to rating this item as being focused on only the most recent session, in contrast to the remainder of the measure for which she rated based on the preceding 2-4 sessions.

***Tasks subscale*.** Codes applied to items from the Tasks subscale encompassed 24.1% (26/108) of the sample; eight participants contributed reflections on items in the Tasks scale.

Item 12 – "I believe the way in which we are working on my problem is correct."

Eighteen codes were applied to utterances regarding this item, which accounted for 17.6% (19/108) of the sample. Six participants contributed codes to this item (see Appendix).

Five participants identified concerns with one or two undefined or vague terms in this question: both the terms 'problem' and 'correct.' These participants identified the word 'correct' as striking them as definitive and narrow, and highlighted their sense that correctness was variable across people and across time. One participant described it in this way: "I don't think there's a correct way of dealing with your problems, I mean there's good ways and bad ways, but there's ways that work, so everybody's different, everybody's going to work their own way" (WAI_32). Two participants further described their beliefs that

[7] Note also her lack of clarity as to who views/collects her completed WAIs. This observation is beyond the scope of the CCS, and thus was not formally captured in these analyses.

their treatments were more complex than simply focusing on a single 'problem. ' They described having been working on multiple areas of concern over their courses of treatment, and thus the word 'problem' was believed to be more narrow than warranted.

Three participants highlighted concerns that this item might lead to desirability bias or be stigmatizing for participants. Two of these participants noted that they were not particularly hopeful at the beginning of sessions but were hesitant to rate this item honestly for fear of hurting their therapists' feelings; one of these participants described this concern about all items that implicitly suggested she would 'change' throughout treatment (items 10 and 12 in particular).

One participant reflected on a number of these concerns about item 12 during his narrative. He noted the following:

> I don't like the word 'correct' because then there's this stigma behind what is normal. Maybe some things that I don't want to do are normal…'Correct' has to be something that needs to be fixed, and I don't think that anybody needs to be fixed…and also the word 'problem' too, I don't know, 'problem' and 'correct' in the same sentence makes you feel as though you're wrong when you came here, it doesn't feel very nice. (WAI_31)

This same client identified that he seldom actually read this item when completing the WAI-SR after sessions, as he reported that his answers for earlier questions essentially dictated his ratings for this item (and item 11); by the time he reached item 11, he was already completing his ratings on autopilot based on his earlier ones by using his typical ratings to complete these final two items.

Item 1 – "As a result of these sessions I am clearer as to how I might be able to change." There were four codes (3.7%; 4/108) applied to this item. Two participants contributed to reflections on this item (see Appendix).

One participant identified this as a yes or no question (a different participant from that mentioned above about previous items), and thus believed the graded rating scale to be unnecessary. This participant also cited concerns about this question when she first encountered it, reflecting that she was not clear how to answer this on first meeting her therapist because she did not develop clarity as to necessary changes until she began to attempt those changes in the world.

Another participant highlighted concerns about desirability bias with this item. She noted her tendency to inflate ratings, particularly in earlier sessions, out of concern for her therapist's feelings. She described this in the following exchange with the interviewer:

> P: It felt like there was pressure, so I felt like a four was, like if no one was looking at this and I filled it out just for my own sake, sometimes I would've probably rated it lower
>
> I: Like what would have been more accurate? Just a ballpark
>
> P: 'Sometimes'
>
> I: So a two?
>
> P: I could have put two depending on how I was feeling, yeah depending on my mood. I know there were points in therapy where I got to a point and I'm like, how effective is this? How much, am I sure that this is allowing me to always change? And so, if it was just for me, I probably would have rated it lower. (WAI_26)

Over the course of treatment, this participant identified leaving sessions confident that her treatment had increased her clarity as to what she needed to do to change. As she highlighted above, though, this was divergent from her early experiences: "In the beginning I had more sessions where I wasn't sure if the sessions were helping or not, or how much they were helping, it wasn't as clear to me" (WAI_26); this apprehension was not reflected in her early ratings.

Item 10 – "I feel the things that I do in therapy will help me to accomplish the changes that I want." Two codes were applied to this item (1.9%; 2/108); these codes came from two participants (see Appendix).

One participant highlighted her dissatisfaction with the rating scale options applied in this item, wishing she were able to provide something in between the existing points: "Sometimes I wish there was a between 'Fairly Often' and '[Very] Often' in terms of the things that we do" (WAI_27).

Another participant noted her tendency to inflate ratings on this item, among others, in early sessions, as she had lower expectations about treatment being effective for her:

> In the beginning I probably bumped up…the ones that were about changing 'cause I wasn't thinking that this was going to help me change, so like #10 though it probably would have been 'Sometimes,' but I probably put like 'Fairly Often' or 'Very Often.' (WAI_25).

In comparing this participant's reflections with her actual WAI-SR ratings during treatment, she did provide ratings of 'Fairly Often' on item 10 for sessions one and three

(the first two sessions rated) with her ratings increasing to 'Very Often' in subsequent sessions.

Item 2 – "What I am doing in therapy gives me new ways of looking at my problem." Only a single code from a single participant (see Appendix) was applied to this item (0.9%; 1/108). It is notable that the other participants who reflected concerns about the word 'problem' as utilized in item 12 did not raise objections to its use in this item.

The single participant's reflections about this item concerned the reference period covered by the measure; her comment was focused on this item as an exemplar of the rating scale reference period as a whole. She identified confusion as to whether she was being asked to reflect on all of her sessions or on the most recent session in particular. She reported her typical rating strategy as follows: "I basically do a quick scan of all my therapy sessions, and try to pinpoint a moment where he hasn't helped me try and find a new way of looking at something, and if that doesn't come up I can say 'Always'"(WAI_29). She thus described having made the decision to reflect on the entirety of treatment in trying to search for explicit memories that represent exceptions to the default idea that her sessions always offered new ways of looking at her problem.

Goals subscale. Goals items contributed the fewest codes to the current project, comprising 18.5% (20/108) of the sample. Five participants contributed to codes reflecting on Goals items.

Item 11 – "_____ (my therapist) and I have established a good understanding of the kind of changes that would be good for me." Seven codes were identified as relevant for this item (6.5%; 7/108); three participants contributed to these codes (see Appendix).

One participant described concerns about the term 'good' as used twice in this item, citing confusion as to how he would be able to gauge whether he and his therapist had in fact established a good understanding of possible changes. He also noted difficulties with the length of the sentence and the repetition of the term 'good.' The same participant, as noted in item 12 above, conveyed his tendency to seldom engage with this question fully because by the time he reached this item, he would essentially start to rate based on his previous answers.

One participant noted her tendency to rate this item based on the previous 2-4 sessions, rather than on the most recent session or the entirety of treatment; she noted this tendency for most of the items on the measure, using this item as her exemplar. Another participant viewed the response options for this item as based on a percentage of agreement, rather than a rating of the frequency with which she agreed with the item's statement, and thus believed the response options to be inappropriate for item 11.

Item 4 – "_____ (my therapist) and I collaborate on settings goals for my therapy." Seven codes were utilized in reference to this item, representing 6.5% (7/108) of the sample. Three participants contributed reflections on this item (see Appendix).

One participant noted a potential desirability bias in this item. She raised this concern about the measure in general, as well as this particular item.

Another participant reflected confusion about the definition of the word 'collaborate' in this item. She described being unclear as to whether collaboration required a totally even distribution of the workload between client and therapist. The participant went on to note her preference for the word 'agree' as used in other items because she found this to be more easily intelligible.

The final participant noted her sense, as described in previous items above, that this item, along with several others (4, 6, 9, 12), is not applicable at each session, and thus she would have preferred not to answer this item after certain sessions: "If this questionnaire is for specific sessions, then those particular ones on certain sessions, if I had a choice not to answer them, I wouldn't" (WAI_27). This excerpt also highlights the participant's lack of certainty as to the reference period covered by the measure.

Item 6 – "_____ (my therapist) and I are working towards mutually agreed upon goals." Six codes were relevant for this item (5.6%; 6/108). All six codes came from a single participant (see Appendix).

This single participant highlighted several areas of focus. She noted that having mutually agreed upon goals was not relevant for her at each session, as she and her therapist did not always identify specific goals on which to work. She thus would have preferred to have a 'not applicable' response option, as she highlighted that this item was not relevant for each completed session.

Item 8 – "_____ (my therapist) and I agree on what is important for me to work on." Participants provided zero specific utterances about this item, and thus zero codes were applied.

Summary. Interviews with the 10 participants from Phase Two incorporated two interview protocols: verbal probing (*n*=5) and think-aloud (*n*=5). Interviews took place with participants during their active completions of the WAI with the intention of identifying item-specific problems. These problems were coded using the Classification Coding Scheme (CCS; Rothgeb et al., 2005). The CCS categorized questionnaire concerns under four broad code families: Comprehension and Communication, Retrieval from

Memory, Judgment and Evaluation, and Response Selection. Each of these code families corresponded to a stage in the four-stage Tourangeau-Rasinski model (Tourangeau & Rasinski, 1988) on which the CCS is based. This model described the processes respondents engage in when completing a survey. These four stages are: 1) Comprehending an item; 2) Retrieving relevant information from memory to answer the item; 3) Evaluating an appropriate response based on the question and memory; and 4) Selecting that appropriate response from the rating options provided.

What follows is a summary of Phase Two results, as grouped according to the type of concern identified by participants. These are thus collated according to the code families of the CCS (no Retrieval from Memory problems were identified in this sample). This is not intended as an exhaustive summary of each item and/or problem code but is presented as a snapshot of salient results. Participants identified a number of concerns regarding most of the items on the WAI-SR. Every item on the measure, other than #8, had at least one concern raised in relation to it.

Codes within the Comprehension and Communication family comprised the majority of the issues participants identified during Phase Two interviews. Of these, issues of Question Content were most frequently noted, particularly regarding what participants identified as vague items and/or vague terms used within items; nine participants identified at least one concern in this vein. The term 'likes' in item 3 ("I believe _____ likes me") was frequently identified as one such term, as was the term 'appreciates' in item 7 ("I feel that _____ appreciates me") and the two words 'problem' and 'correct' used in item 12 ("I believe the way in which we are working on my problem is correct"). Five participants also highlighted occasional concerns with Question Structure. Specifically, they

noted concern with an item making an erroneous assumption, such as that a participant would have done something about which their therapist might have disapproved, which is implicitly assumed in item 9 ("I believe my therapist still cares about me even when I do something he/she does not approve of"). Finally, two participants highlighted ambiguity in the period queried by the items in the measure (i.e., whether they were being asked to rate based on only the recent session or based on multiple sessions/all of treatment).

The next most frequently applied code family was Judgment and Evaluation. Specifically, participants identified concerns about a possible desirability bias which is considered a concern when evaluating an 'appropriate' response. Six participants highlighted possible issues with desirability bias across several items, such as in items 3 ("I believe _____ likes me") and 12 ("I believe the way in which we are working on my problem is correct"). They indicated that such items called for a high rating.

Response Selection issues were the next most-frequently identified by participants. Four participants believed that the scale was missing possible response categories, such as a 'Not Applicable' rating option, which could help to obviate the concern about erroneous assumptions in items like #9 described above; some participants noted they either wrote 'N/A' on their own or simply skipped such items. Two participants conveyed their beliefs that certain items required a dichotomous response and were not appropriately rated on a graded scale, such as believing respect to be a dichotomous state as queried by item 5 ("_____ and I respect each other).

Phase Three Results

Overview. Phase Three was intended to provide a triangulated perspective of the findings produced in Phases One and Two. These reflections on points of convergence in

the findings were identified to help synthesize and elaborate the results into a larger whole. It utilized an intuitive triangulation procedure (Farmer et al., 2006; Meijer et al., 2002) and sought to present findings in multiple forms, seeking to approach crystallization in the vein of Ellingson (2014). To further enhance the diversity in the representation of clients' perspectives on the WAI, an intensive case analysis of one Phase Two participant's think-aloud interview is presented to elucidate points of similarity and difference across the phases.

Triangulating findings. Following Farmer et al. (2006), the points of convergence, divergence, and thematic silence across both phases were identified. Considering the highest degree of convergence occurred between the Phase One main category *Making Sense* and the Phase Two findings generally, these are described first. These item- and measure-specific points of convergence speak directly to the second research question of this study; namely, what concerns do clients identify when completing the WAI-SR items? Other areas of convergence are then discussed. There was no clear evidence of divergence identified across phases, though one possibility is noted. There were a number of points of thematic silence between phases, though this was not entirely surprising considering the differing intentions and foci of each phase.

There was considerable overlap across the Phase Two findings and the Phase One main category *Making Sense*. *Making Sense* captured participants' understandings of the items and the scale, which were explicitly queried throughout Phase Two. The *Making Sense* subcategories overlapped clearly with several of the code categories found in the Classification Coding Scheme (CCS) used to code participant responses during Phase Two. The *Making Sense* subcategory *Figuring out what the question is asking* comprised

reflections similar to those captured by the CCS code families focused on Questions Content and Question Structure. This subcategory emphasized participants' questions or difficulties with the items on the WAI. In terms of specific reflections, items identified as challenging in Phase One converged with those found in Phase Two, particularly as regards participants' reflections about the ambiguity and complexity of the term 'likes' in item 3 and questions about its application and relevance in the therapeutic context. Nine of the 10 participants in Phase Two identified at least one concern (range = 4-28) about items on the WAI. Three quarters of these concerns (77/108) related to difficulties with comprehending item content and structure; the bulk of those difficulties (53/77) reflected confusion about item content and the meaning of questions and/or terms used within questions.

Findings from *Figuring out what the question is asking* also connected with concerns identified by the CCS' Erroneous Assumption code as applied to item 9 and, more specifically, the assumption that participants would have done something about which their therapists might have disapproved; five out of 10 Phase Two participants identified item 9 as problematic for this reason. When faced with an item querying an inapplicable situation, participants across both phases elaborated on the various avenues through which they had dealt with it: 1) They inferred how their therapists would likely act *if* such a situation had occurred and then provided the corresponding rating; 2) They skipped the item entirely or added their own response option of 'N/A;' or 3) They answered at their 'baseline' rating (e.g., 3/Fairly Often), as they were unsure what an appropriate rating would be.

Another *Making Sense* subcategory, *What does 'Sometimes' appreciating someone mean?,* accounts for concerns similar to those captured in the CCS family Response

Selection. Specifically, participants in both phases conveyed the preference to have a 'Not Applicable' response option added to the scale to help alleviate the need to provide a numerical/descriptive response when situations were not relevant for the participant. Examples of inapplicable situations noted by participants included the one described above referenced in item 9, as well as the need to answer items about explicitly setting goals after the first few sessions. These goals-related items were seen by some participants as only becoming relevant after several sessions had passed and therapists and participants had developed a shared sense of each other (as described in the *Making Sense* subcategory *Figuring out what the questions is asking*).

A final area of convergence with *Making Sense* and Phase Two findings related to the reference period of the WAI. Participants across both phases noted ambiguity with respect to the period queried by the WAI. Confronting this ambiguity, Phase One participants described several routes through which they provided meaningful ratings (as explored in the *Making Sense* subcategory, *Sometimes I'll rate until now, sometimes just this session* and the CCS code family Reference Period): 1) Basing ratings only on the most recent session; 2) Basing ratings on their experience across all sessions; or 3) Using a hybrid approach that included basing a rating on the most recent session when something relevant for the item had occurred in that session and otherwise basing a rating on experiences across treatment. Many participants identified using the latter hybrid approach.

The Phase One subcategory *If we've had that conversation…* comprised reflections related to the CCS code family Retrieval from Memory, the only family from which no codes were applied during Phase Two analyses. The Retrieval from Memory family implies, based on the Tourangeau-Rasinski model and in line with its cognitive emphasis, that a

respondent must be able to highlight a relevant, explicit memory in order to provide a reasonable answer on a measure. *If we've had that conversation…* challenges this emphasis on explicit cognition and highlights the fact that respondents take different routes to finding relevant information when answering an item. As identified in this subcategory, participants utilized explicit memories, intuitive/emotional reactions, or both as anchors from which to base their ratings. The same individual would often use a combination of the sources of information when addressing different items, approaches that may change depending on the day and the item. In some ways, participants described retrieving implicit attitudes towards therapists or treatment when offering ratings and not necessarily explicit memories.

Another area of overlap between phases included reflections about the possibility of a desirability bias at work within the measure, which was identified by the CCS Desirability Bias code and the *Wrestling With Dissonance* subcategory *I wanted to be honest.* A minority of participants from Phase One identified intentionally inflating their ratings due to worries about their therapists' or supervisors' reactions to low ratings while the majority who reflected on this theme indicated that they attempted to provide honest feedback, even if honesty caused discomfort. A majority of participants in Phase Two (6/10) identified the *possibility* of a desirability bias leading to inflated ratings on certain items such as item 3 ("I believe _____ likes me") and 12 ("I believe the way in which we are working on my problem is correct"), but these participants did not necessarily endorse this tendency themselves. They thus recognized that some elements of the WAI *might* induce someone to provide a response that was more favourable than accurate, yet did not identify having done so themselves.

These reflections on social desirability may also represent a point of divergence or silence between phases, though which remains unclear. A majority of participants in Phase Two stated that certain items could be more difficult to answer honestly, such as items 3 and 12. As noted, this is in contrast to the majority of participants from Phase One who denied (consciously) inflating their ratings. It is notable that in Phase Two, participants were speaking more in the abstract (i.e., one *might* be more likely to inflate ratings), whereas those in Phase One were being asked explicitly if they themselves had engaged in ratings inflation. There were no other evident points of divergence across phases.

In light of the differences in the goals and methods between Phases One and Two, there were a number of reflections from one phase that did not have correlates in the other phase (i.e., there was thematic silence across the phases, in the sense of the triangulation procedure utilized by Farmer et al., 2006). Phase Two interviews sought explicitly to identify and collate areas of concern with regard to the measure itself, and thus there were a number of categories and subcategories developed during Phase One analyses that had no correlates in Phase Two. The main categories of *Empathic Leading* and *Working Through*, which pertained to helpful relational and intervention elements, for example, were not coded during Phase Two. There were likewise item-specific reflections developed in Phase Two that had no Phase One parallels, or at least no parallels with the depth of those identified in Phase Two. While there was considerable overlap between the Phase One *Making Sense* main category and Phase Two findings, as described above, the level of item detail evidenced in Phase Two was unique to that phase. These areas of silence were not surprising, as the intentions and questions influencing the design and implementation of each phase were distinct.

Single case study. The ultimate purpose of this study as a whole was to examine how clients understand and rate their experiences on the WAI-SR. The analysis and synthesis of interview data represents, of course, a key means with which to help understand those processes to a greater degree. The analytic process intentionally distances the individual participant from the results, aggregating data across the samples to provide general impressions. This has great merit. This process, however, also distances us from participants' actual, dynamic rating processes. Knowing this, an interview with a participant from Phase Two (think-aloud protocol) is presented below as a complementary perspective through which to understand clients' rating procedures. It is hoped this interview will complement the synthesized findings described above and help re-inject the client into this research process. In the vein of crystallization (Ellingson, 2014), this interview is presented as another facet through which to experience the rating phenomenon and one I believe to be important to afford deeper access to a client's unfiltered rating procedure (insofar as any rating procedure can be unfiltered in a verbal, recorded, and scheduled interview with a stranger). This think-aloud interview is also presented to supplement the Phase Two interview data, as some participants' reflections inevitably fell outside the scope of the predetermined CCS code categories, and thus were not captured within Phase Two.

For the sake of brevity, only the initial narrative think-aloud portion of this interview has been reproduced herein, as this is the only component in which the participant offered (relatively) unfiltered access to her rating processes. The think-aloud portion of this interview provides insight into the client's spontaneous item-by-item process while completing the WAI-SR. This think-aloud portion of the interview is first

presented in its entirety. A brief analysis of her responses to each item as compared with the general findings of this study follows thereafter. The WAI-SR item numbers have been added and minor grammatical changes made to improve ease of reading. Note that the items themselves are embedded within the interview transcript and analysis (after the added item numbers), as participants were asked to read each item aloud as they worked through the measure and narrated their rating processes.

This interview was completed with participant WAI_23, a 21-year-old female undergraduate student at York and the first participant interviewed during Phase Two. The interview took place just after her final session, the 14th in her course of therapy. She had a mean WAI-SR item rating of 4.7/5 for the entirety of her treatment and completed seven WAI-SRs during treatment, one after every second session. Thus, the WAI-SR completed in her research interview represented her eighth completion of the measure. Of note from past completions of the measure acquired from her file, the WAI-SR completed following her first session showed that she wrote '?' beside items 3, 7, and 9, even though she still completed each item (rating a 3, 4, and a double rating of 3 and 5 for each item, respectively). If we exclude her double rating for item 9 (since there is no definitive rating here), her first session mean WAI score was 4.25/5, her third session WAI mean was 4.33/5, and every completion thereafter was scored with either a mean item rating of 4.92/5 or 5/5. From a quantitative perspective, she evinced a strongly positive alliance. As we will see, her narrative conveyed in interview echoed this positivity in her therapeutic connection, but it also suggested a more nuanced view and doubt about her ratings than a strictly quantitative analysis would represent.

Single case: Think-aloud interview.

[Item 1] "As a result of these sessions, I am clearer as to how I might be able to change", it's kind of a weird question to ask on the first day, like the first time that I was asked it, but I feel like I won't actually be clear until I put it into practice, and yeah I've chosen five because oftentimes I feel like I've come away with something and therefore I'm clearer than I was when I first arrived, I mean I wouldn't know whether or not to rate it a four or a five or a three, it's just kind of a yes or a no for me. **[Item 2]** "What I'm doing in therapy gives me new ways of looking at my problem," similar to question one, but I'd say maybe four or five because again I always feel like I'm going away with something even if it's not direct, it's just the fact that now my mind is going as to what I might want to do even if I haven't, it's not direct if that makes sense, it's like I'm just now thinking about what I might change or how I put that into practice. **[Item 3]** "I believe [my therapist] likes me," kind of a weird question, but I just answer 'Always' because, or usually answer 'Always' because I can't, I wouldn't, I mean not that I don't think it's relevant because obviously it is, but he's going to be, he's going to come off as, what am I trying to say? In a way in therapy you're playing roles and in that role a therapist is supposed to like you, so regardless he's going to like me. **[Item 4]** "[My therapist] and I collaborate on setting goals for my therapy," I'd say 'Always' again because I think that the way that we work through things or talk about it, there's a bit of a back and forth, and I feel that he generates his ideas based off of whatever I'm saying, so there's collaboration. **[Item 5]** "[My therapist] and I respect each other," again 'Always' because there hasn't been any instances when I felt like there was some

disrespectful vibe so that's that, and I guess similar to the one about whether or not he likes me, there's just that, not a natural respect but that's just the nature of the relationship. **[Item 6]** "[My therapist] and I are working towards mutually agreed upon goals," also sounds similar to number four, yeah I also generally say 'Always' because I don't feel like it's something that's being forced on me or he's imposing his ideas, he's kind of working off of whatever I say, and then sometimes I feel that the goals that are developed are my own that are just being put on paper. **[Item 7]** "I feel that [my therapist] appreciates me," it's kind of weird because he's providing the sessions for me, so I still say 'Always,' but I don't really understand that, but I guess again in his demeanour he's appreciative in that, well if you want to pair 'appreciate' with 'respect' and it kind of plays out similarly, yeah I think that's a strange question. **[Item 8]** "[My therapist] and I agree on what is important for me to work on," well yes but in that sense I feel that he didn't really guide, again it would be me telling him what I felt needed to be worked on. **[Item 9]** "I feel [my therapist] cares about me even when I do things he does not approve of," I guess I would say five ['Always'], but I wouldn't think that because I wasn't always open to share everything, not that I don't think that he would approve of, I think he would but I was still hesitant to share certain things. **[Item 10]** "I feel the things I do in therapy will help me to accomplish the changes that I want," 'Always' but at times I felt well maybe I have something due or something, so maybe that would be better for me to do at this time, to work on those things but I think I would say 'Always' because in therapy sometimes I talk about things I wouldn't necessarily think of outside of it and it has helped me a lot. **[Item 11]** "[My therapist] and I have

established a good understanding of the kinds of changes that would be good for me", I would probably put 'Always,' but at this point I'm not really thinking about the question, I'm just going off of what I've previously done and it sounds like it would go number five. **[Item 12]** "I believe the way that we are working with my problem is correct," I pretty much always say 'Always' but in my head I'm thinking there's the possibility that there's other things I'm missing out on, I don't really know, I hope not

Single case analysis. This participant described a number of topics that were reported by other participants. Overall, she conveyed her experience in therapy as a positive one, identifying her therapist and treatment as helpful. The vast majority of participants in both phases of the present study shared this perspective. During follow-up queries later in the interview, she reflected on her experience completing this measure generally: "I guess every time I fill it out it's a strange feeling. I feel a bit awkward with this particular form." She elaborated on this discomfort, noting that items referring to 'liking' or 'respect' with her therapist (items 3 and 5, respectively) were what made her feel awkward. Her responses to these items are described in greater detail below.

The participant identified a number of item-specific issues that shared similarities with those from other participants. Item-by-item analyses of this participant's think-aloud narrative follow, with additions from her later responses to follow-up queries as relevant. Her narrative for each item is presented again below to allow for ease of comparison between her narrative and the analyses that follow.

[Item 1] "As a result of these sessions, I am clearer as to how I might be able to change", it's kind of a weird question to ask on the first day, like the first time that I

> was asked it, but I feel like I won't actually be clear until I put it into practice, and yeah I've chosen five because oftentimes I feel like I've come away with something and therefore I'm clearer than I was when I first arrived, I mean I wouldn't know whether or not to rate it a four or a five or a three, it's just kind of a yes or a no for me.

In her description of item 1, the participant highlighted key reflections that were also relevant for a number of other participants across both Phases One and Two. Firstly, she noted this to be a "weird" question when responding to it after the first session, due to the fact that she had yet to actually attempt to make any treatment changes in her life, which was consistent with some participants' perspectives in Phase One. She conveyed her sense that clarity about making changes was predicated on beginning to implement those changes in her life, which she had not done after the first session. During follow-up questions, she identified her inference that she rated this item highly on the first day (which she did, rating a 4/Very Often that first time) because, "I felt like there was potential for the future that I would feel good"; in short, her expectations for treatment effectiveness led to this first session rating. This reflection about initial expectations that therapy would be helpful in the absence of early treatment gains was echoed in Phase One findings.

The participant also noted her sense that this item necessitated a dichotomous rating (either a yes or a no) rather than a rating on a graded scale. The reflection that at least some items were consistent with a dichotomous rather than dimensional scale was evident in findings from both phases. Given that she was presented with a graded scale, the participant selected 5/Always because she "oftentimes" had greater clarity than when she began treatment. She thus evinced some discrepancy between the scale descriptor of

'Always' and her personal view that she 'oftentimes' came away from sessions with greater clarity, thus evincing the Phase One strategy noted by some participants whereby they personalized certain points of the scale.

> **[Item 2]** "What I'm doing in therapy gives me new ways of looking at my problem," similar to question one, but I'd say maybe 4 or 5 because again I always feel like I'm going away with something even if it's not direct, it's just the fact that now my mind is going as to what I might want to do even if I haven't, it's not direct if that makes sense, it's like I'm just now thinking about what I might change or how I put that into practice.

In describing item 2, she did not identify concerns about the item's structure and intelligibility. She did note her awareness of the similarity between items 1 and 2 (both are included on the Tasks subscale) and described her approach to selecting a "4 or 5" rating as due to the fact that she "always" left sessions with a sense of the kinds of changes she might like to undertake, even if she had yet to attempt to implement these in her life.

> **[Item 3]** "I believe [my therapist] likes me," kind of a weird question, but I just answer 'Always' because, or usually answer 'Always' because I can't, I wouldn't, I mean not that I don't think it's relevant because obviously it is, but he's going to be, he's going to come off as, what am I trying to say? In a way in therapy you're playing roles and in that role a therapist is supposed to like you, so regardless he's going to like me.

She described item 3 as "weird." The perspective that the term 'likes' was somehow ambiguous, irrelevant, or difficult to rate was evidenced across both phases. This participant conveyed surprise here because she believed it to be a given that her therapist

would convey that he liked her, as she perceived this to be axiomatic of the therapist's professional role. She thus 'Always' believed he liked her, but she appeared to question the relevance of this item as providing meaningful information to others.

> **[Item 4]** "[My therapist] and I collaborate on setting goals for my therapy," I'd say 'Always' again because I think that the way that we work through things or talk about it, there's a bit of a back and forth, and I feel that he generates his ideas based off of whatever I'm saying, so there's collaboration.

The participant did not highlight concerns about the interpretation or relevance of this question, and thus implicitly suggested it to be understandable and pertinent. She conveyed the basis for her 5/Always rating as being her sense that her therapist offers a "back and forth" with her and grounded his suggestions for her in what she had already shared with him.

> **[Item 5]** "[My therapist] and I respect each other," again 'Always' because there hasn't been any instances when I felt like there was some disrespectful vibe so that's that, and I guess similar to the one about whether or not he likes me, there's just that, not a natural respect but that's just the nature of the relationship

She did not report concerns about this item's structure. She did report utilizing a strategy similar to that elaborated by Phase One participants, namely, basing her 5/Always rating on the *absence* of disrespectful interactions with her therapist. In this reflection, she exemplified the approach described by Phase One participants in which they anchored their ratings to the 'Always' rating by default, assessed the accuracy of this presumed rating by scanning their memories for exceptions, and then judged the rating to be accurate based on the absence of exceptions.

This participant also highlighted the similarities between this item and item 3 (both are Bond subscale items) and described this item's referent of mutual respect as again being axiomatic in a therapeutic context, as was the case for 'liking' referenced in item 3. She thus appeared to be unsure as to the relevance of her ratings on this item.

> **[Item 6]** "[My therapist] and I are working towards mutually agreed upon goals," also sounds similar to number four, yeah I also generally say 'Always' because I don't feel like it's something that's being forced on me or he's imposing his ideas, he's kind of working off of whatever I say, and then sometimes I feel that the goals that are developed are my own that are just being put on paper.

The participant again did not identify concerns about this item's structure or meaning. She did note the similarity to item 4 (another Goals item) and, as with item 4, described her sense that her therapist based his suggestions on what she was identifying as important, rather than "imposing" his views onto her.

> **[Item 7]** "I feel that [my therapist] appreciates me," it's kind of weird because he's providing the sessions for me, so I still say 'Always' but I don't really understand that, but I guess again in his demeanour he's appreciative in that, well if you want to pair 'appreciate' with 'respect' and it kind of plays out similarly, yeah I think that's a strange question.

As with her account regarding item 3, the participant identified item 7 as "weird" and difficult to understand. This response was consistent with reflections from participants in both phases but was particularly evident in the findings from Phase Two; this was the most frequently cited WAI-SR item of concern in Phase Two. She reported confusion as to the meaning of the term 'appreciates' being used in this context. During follow-up queries, she

elaborated on this confusion, explaining her therapist to be an "appreciative person, like he's a good person, if that's what you're asking."

The participant also implicitly described a process similar to that identified by participants in Phase One, whereby, when confronted with a confusing term like 'appreciates,' she looked to another more understandable term, like 'respects,' as a guide from which to respond to this item. She later described her rating choice further: "I give a 5 because it sounds good…I would think of ['appreciates'] as like 'respect,' I would pair those together."

> **[Item 8]** "[My therapist] and I agree on what is important for me to work on," well yes but in that sense I feel that he didn't really guide, again it would be me telling him what I felt needed to be worked on.

She again did not highlight concerns with this item and reiterated the basis on which she selected her rating. Consistent with her descriptions in items 4 and 6 (others Goals items), she based her rating on the degree to which she helped her therapist to identify salient goals for therapy.

> **[Item 9]** "I feel [my therapist] cares about me even when I do things he does not approve of," I guess I would say five, but I wouldn't think that because I wasn't always open to share everything, not that I don't think that he would approve of, I think he would but I was still hesitant to share certain things

The participant did not identify concerns with the interpretation or meaning of this item, which was consistent with findings from both phases. She did convey her belief that, though she had not been completely open with her therapist, she had a sense that he would not have disapproved if she had done so. During follow-up, she highlighted her belief that,

> the purpose I got out of the question is that it's questioning your openness. It doesn't say that, but even though I would not give myself a high rating on that, I feel like if I wanted to share those things, that he would approve.

This is an example of the participant interpreting the item perhaps slightly differently than intended, focusing more on her own openness (or lack thereof) rather than her therapist's consistent caring attitude in light of perceived disapproval.

The participant's response also highlighted her belief that the item was not entirely relevant for her, which was likewise consistent with findings across both phases. Numerous participants in both phases expressed difficulties identifying memories in which they had shared something that the therapist might disapprove of, thus indicating that the item made an erroneous assumption about the therapy encounter. This participant in particular conveyed her belief that this was not entirely relevant because she had not given her therapist sufficient reasons to disapprove insofar as she had not been totally open about certain things. In light of this lack of context, this participant described a process consistent with one outlined in Phase One findings, whereby she made an assumption based on her therapist's attitude in general that he would not have disapproved if a relevant situation had arisen.

> **[Item 10]** "I feel the things I do in therapy will help me to accomplish the changes that I want," 'Always' but at times I felt well maybe I have something due or something, so maybe that would be better for me to do at this time, to work on those things but I think I would say 'Always' because in therapy sometimes I talk about things I wouldn't necessarily think of outside of it and it has helped me a lot

The participant did not note concerns about the intelligibility of this item. She here introduced a separate issue that, at this point in the measure, she often started to think about leaving her session and getting back into the rest of her life, and thus considered the final items less carefully than the earlier ones. This tendency became more pronounced when she described items 11 and 12.

> **[Item 11]** "[My therapist] and I have established a good understanding of the kinds of changes that would be good for me", I would probably put 'Always' but at this point I'm not really thinking about the question, I'm just going off of what I've previously done and it sounds like it would go number five.

The participant extended her reflection about a tendency to expend less energy on this penultimate item. Her reasoning for this item, as she described here and during follow-up queries, is that item 11 appeared similar to items 4 and 6 (all Goals items) and she answered this item "based off the other ones…by that point I'm pretty sure in myself that yes there's an understanding, so I just give a 5." She again did not highlight challenges understanding this item.

> **[Item 12]** "I believe the way that we are working with my problem is correct," I pretty much always say 'Always' but in my head I'm thinking there's the possibility that there's other things I'm missing out on, I don't really know, I hope not

There were no reported problems interpreting this item or its terms. She highlighted some degree of concern about being unaware of other "correct" ways to address her "problem." This doubt was not captured in her ratings.

Single case conclusions. This participant's think-aloud interview provided a clearer view of the client rating process than those offered in the previous phases. In her interview

account, the participant above described understanding most of the items quite easily and grappling with the meanings or relevance of a few of the others. She demonstrated an awareness of the similarities across some of the items within the same subscales and used this awareness to help provide a rating for an ambiguous item (identifying similarities between terms 'appreciates' and 'respects' in items 7 and 5, respectively). The active nature of her engagement with some elements of the measure was clear. As she had been a veteran WAI-SR rater at the time of the interview, one wonders how much more active her process in making meaning and responding to the scale would have been had she been interviewed upon first encountering the measure.

Discussion

Study Overview

This study employed two types of intensive analytic methods across two primary phases of inquiry, and a third secondary phase, to help develop a deeper understanding of clients' perspectives of and approaches to rating the Working Alliance Inventory (WAI; Horvath & Greenberg, 1989), and the WAI-Short Revised (Hatcher & Gillaspy, 2006) in particular. In the first phase, Grounded Theory (Corbin & Strauss, 2008) was used to qualitatively analyze transcripts generated from semi-structured interviews conducted with 12 clients in ongoing psychotherapy. These analyses synthesized clients' general perspectives on the rating process and the measure itself. Phase One yielded a core category of **Rolling on the Right Track** and seven main categories that will be described in greater detail throughout the section that follows. The second phase utilized Cognitive Interviewing methods, specifically verbal probing and think-aloud procedures (Priede & Farrall, 2011; Willis, 2004), and a content analytic coding framework, the Classification Coding System (CCS; Rothgeb et al., 2005) to analyze a further 10 additional interviews with clients engaged in psychotherapy. These interviews produced item-specific findings derived using the CCS about clients' understandings of the WAI-SR, which will also be described below. The third phase was a secondary analysis that assessed the convergence of findings across both phases and offered an in-depth presentation of one participant's think-aloud interview.

There were three broad research questions guiding the present work, overlapping with the three phases described above: 1) What are clients aware of when they complete the WAI?; 2) What item and scale concerns do clients identify when completing the WAI?;

and 3) How do clients understand and engage with the WAI overall? The discussion that follows will elaborate on key findings that followed from these research questions and will explore them along with other findings presented in the literature.

Key Findings

The core category, **Rolling on the Right Track**, derived from the Grounded Theory analyses of participant interviews in Phase One, fits nicely with Hatcher's (2010) definition of the alliance:

> Across different therapeutic approaches, a good alliance means that patients and therapists are working well together towards the goals of therapy. Good work is expected to be purposeful and collaborative. Thus, the alliance is a way to think about how the patient and therapist are working together. (p. 8)

This definition highlights the directed, "purposeful" nature of the alliance, while also stressing the importance of collaboration in therapeutic work. The core category **Rolling on the Right Track** likewise emphasizes clients' views that the WAI is a measure focused on assessing how clients and therapists are moving together towards shared goals. This view presents the alliance as a measure of the degree to which clients believe in and are engaged with the interventions and goals of their particular courses of therapy. Participants in Phase One identified this element of the alliance as being central to their conceptualizations of the purpose and focus of the WAI, as well as to their beliefs about what the WAI sought to measure (i.e., the alliance itself). Participants described the WAI as a means with which to assess and convey the alignment with their therapists in helping to produce therapeutic changes.

The main categories developed from participants' Phase One interviews about their experiences completing the WAI encompassed two key facets, namely, participants' attempts at *Translating* their experiences in the therapeutic work to answer WAI items, as well as their actual experiences of *Relating Together* with their therapists. The *Translating* main categories include *Making Sense, Anchoring Ratings, Wrestling with Dissonance, Impact Over Time,* and *It's More Complicated*. The *Relating Together* main categories comprise *Empathic Leading* and *Working Through.*

Phase Two yielded item-specific results in which participants highlighted areas of ambiguity or concern with the WAI-SR. Concerns about certain terms or item phrasing were the most frequently raised, particularly with the term 'likes' in item 3 and 'appreciates' in item 7. Concerns regarding item 9 referenced an erroneous assumption that clients had done something of which therapists would have disapproved. Participants also highlighted possible concerns with social desirability biases across some items, as well as indicating a preference that the scale includes a 'Not Applicable' response option.

Participants in both phases of this study made active attempts towards helping themselves understand the meaning of items, personalizing the scale, and providing relevant and meaningful responses. The literature on survey responses consistently highlights the active, interpretive nature of questionnaire responding (Galasinski & Kozlowska, 2010; Rosenbaum & Valsiner, 2011; Schwarz, 1999, 2007). This literature contends that self-report measures are not objective instruments that provide results that are wholly and accurately aggregated, but instead access fluid, subjective, and dynamic processes that are personal to the respondents completing

the measure. These active processes involved in survey completion were exemplified and underscored throughout Phases One and Two in the present study.

Alliance Development

The Phase One main category *Working Through* provides support for Xu and Tracey's (2015) reciprocal alliance model in which they posit that early alliance ratings predict symptom change, which in turn predicts subsequent alliance ratings, etc. Echoed in *Working Through*, participants describe the importance of feeling some initial degree of comfort and trust with their therapists (a Bond element of the alliance). This bond allows them to open up and to work towards and experience change, which in turn increases confidence in their therapists' abilities and leads to increases in ratings for (at least) goals-related WAI items. As such, participants' experiences of nascent alliances and treatment gains develop together in a positive feedback loop. *Working Through* likewise supports the 'emotion-exploration spiral' posited by Fitzpatrick, Janzen, Chamodraka and Park (2006), whereby a strong positive emotional connection (Bond) allows clients to feel sufficiently comfortable in therapy to be open and undertake productive exploration (Tasks-Goals of treatment) of salient personal stories and key concerns. Fitzpatrick and colleagues presented findings that suggest the reciprocal and reinforcing relationship between relational/bond factors and intervention/task-goal factors.

In addition, one of *Working Through*'s subcategories, *I feel it every week*, further connects with the findings of Hatcher and Barend's (1996) alliance factor dubbed 'Confident Collaboration.' *I feel it every week* highlighted participants' faith that their work being undertaken in therapy was helpful, purposeful and had

resulted in some degree of change in their lives. Hatcher and Barends (1996) and Hatcher (2010) outlined their views that this conceptualization of the alliance as involving a purposive, productive element (emphasized by Greenson (1967) and Bordin (1979)) had at times been lost in favour of a view of the alliance as more abstractly relational and echoing more of a focus on the 'real' relationship (Greenson, 1967) in therapy. Participants in this study emphasized their beliefs that this purposive element is indeed central to how they perceived therapists and how they conveyed those perceptions on the WAI. They demonstrated an awareness that the intention of the WAI was not solely to gauge the strength of the relationship in isolation, but rather to assess how effective relational and treatment factors are in helping clients meet treatment goals.

These findings also speak to Bachelor's (1995) research regarding clients' perspectives of the alliance in general. She identified that most of her participants conceived of the alliance as nurturant, insight-oriented, or collaborative. These types of alliances, particularly nurturant, share elements with subcategories posited in the *Relating Together* main categories, insofar as current participants' descriptions of therapist non-judgment, trust and empathic listening were identified frequently as integral for making changes over treatment. Bachelor's insight-oriented alliances emphasize the acquisition of heightened self-understanding over therapy. This type of alliance connects with participant reflection in the *Lead them to the right way* subcategory, wherein therapist guidance and suggestions were relayed as being integral.

Relating Together categories also resonate with MacFarlane, Anderson and McClintock's (2015) findings regarding clients' perspectives on formative elements in early alliance. The subcategories *She never disapproved* and *Working positively* share similarities with their findings of client-rated importance of therapists' supportive activities (e.g., offering validation or hope). *It takes time to build* echoes MacFarlane et al.'s (2015) finding that clients are often apprehensive for a variety of reasons when beginning treatment, and, like *It takes time to build*, MacFarlane and colleagues note that therapist supportive behaviours, such as those identified in *She never disapproved* and *Working positively,* can serve to reduce clients' apprehension. *Lead them to the right way* offers connections with MacFarlane and colleagues' clients' appreciation for technical activities undertaken by therapists.

Consistent with the quantitative findings presented in Hatcher and Gillaspy (2006) about the near-perfect correlation between scores on the Goals and Tasks subscales from the original WAI, three Phase One participants appeared to explicitly conflate items between these scales, as described in the subcategory *I don't know the right answer*. Some participants stated their belief that the WAI essentially represented two subscales, one pertaining to relational, emotional elements between themselves and therapists (seemingly corresponding to the Bond items) and a second subscale emphasizing the practical work completed in treatment (seemingly a combination of the Goals and Tasks items). This suggests that the differentiation of the alliance into three subscales as conveyed through the items presented on the WAI-SR may lack some degree of face validity among participants in this study.

Participants' reflections are also in line with Mallinckrodt and Tekie's (2016) conclusions. These authors, among other goals, analyzed WAI data, and, like Hatcher and Gillaspy (2006) and Tracey and Kokotovic (1989) before them, found a lack of support for a three-factor (Bond, Tasks, Goals) structure as a result of the very high correlations between Goals and Tasks subscales. Mallinckrodt & Tekie (2016), building on Hatcher and Gillaspy's (2006) work, hypothesize that, "Perhaps clients think of what the WAI Goals and Tasks items tap as a single 'activities of counseling' construct" (p. 714). The findings from Phase One of the present work in which participants conflated the two provides some support for their hypothesis.

As a result of the lack of support for a three-factor model, Mallinckrodt and Tekie (2016) presented a new alliance measure based on the WAI, the Brief Alliance Inventory (BAI). The BAI is comprised of two subscales: a Bond scale and a Goals-Tasks hybrid scale. In this measure, they also recommend eliminating the frequency estimates of the qualitative descriptors in the scale itself, citing findings that indicate a lack of respondent ability to discriminate among certain scale points. Citing research undertaken by Bocklisch, Bocklisch and Krems (2012) they posit that respondents have considerable difficulty differentiating frequency scale points equidistantly, which leads to concerns of treating such data with parametric statistical tests. The participants in both phases echoed this finding, and reported having difficulties understanding some of the scale points as written. Some also described modifying and personalizing the scale descriptors, further highlighting the challenges of assuming that all respondents understand scales in the same way.

WAI-SR Rating Process

Rosenbaum and Valsiner (2011) highlight several critical observations about the subjectivity and inconsistency inherent in survey completion. In general, they argue that scales take what are by definition subjective, dynamic processes and develop reified, static representations of them. This perspective was put forth by some Phase One participants in the present study, described in *Just a number and a phrase*, who noted that the WAI is by its nature a reductive yet expedient instrument. This concern echoes Stiles and Goldsmith (2010) when they note that reducing the complexity of therapeutic relationships to the broad evaluative construct of the alliance necessarily loses considerable nuance. The findings of the present study underscore the active and fluid nature of participants' response processes at work when reflecting on treatment and providing ratings. Explicating these dynamic response processes relevant for WAI respondents is a primary contribution of this study.

Schwarz (2007) distinguishes between survey items that query attitudes versus those that query explicit behaviours. In the case of the WAI, each item "was designed to capture a feeling, sensation, or attitude in the client's field of awareness" (Horvath, 1981, p. 41), and thus seeks to query client attitudes towards therapists and treatment rather than asking respondents to identify specific behavioural occurrences. When an item queries an attitude, respondents either access a previously formed attitude or derive one on the spot (Schwarz, 2007). Phase One participants in this study stated that when they first encountered the measure, they needed to derive or explicate their attitudes towards therapists/treatment, and thus

they engaged in a longer and more deliberated process to make sense of the measure and to understand how they felt about the therapeutic work. What seems to change over time when rating the WAI is that participants develop a more stable and predictable understanding of the measure itself and likewise develop a more stable attitude towards their therapists and treatment. The change over time can be more readily accessed upon subsequent administrations of the WAI. This process also accounts for the fact that a majority of participants used behavioural information if it had occurred in the most recent session (e.g., if the therapist had appeared to judge the participant negatively), though they referred to their general impressions and attitudes to treatment overall if there was no relevant, recent behavioural referent. This perspective of alliance ratings as measures of attitudes to therapists also connects with some participants' reflections in both phases that certain items, particularly Bond items, were more amenable to a dichotomous scale rather than a graded one (e.g., either my therapist likes me or she does not); the assessment of frequency, while relevant for behavioural items, is less relevant for attitudinal ones.

Participants described attending to both the literal meanings of the words and punctuation, as well as inferring the pragmatic meaning (i.e., using the meaning and context of an item to infer why an attitude or behaviour is being queried) (Schwarz, 1999, 2007) to understand items on the WAI. These reported processes for inferring pragmatic meaning became particularly apparent where participants in both phases identified confusion about the meaning of a term or item. For example, when some participants in both phases encountered the terms 'appreciates' and

'likes' and found them to be confusing, they would infer what was being asked of them, identify a salient synonym, and use this seemingly comparable word (e.g., 'respects') to help make the item more intelligible. While making inferences as to pragmatic meaning, respondents used the context of the item, the measure as a whole, the setting in which it is being asked, what/who is being queried, etc.; participants do not complete measures in a vacuum (Rosenbaum & Valsiner, 2011; Schwarz, 2007). This use of pragmatic meaning further underscores participants' active, dynamic engagement with the WAI.

In an experiment investigating questionnaire context effects, Galesic and Tourangeau (2007) found that respondents assessing the frequency of behaviours on a 'sexual harassment' measure were more likely to identify those behaviours as indicative of sexual harassment, rate them as more problematic, and occurring with greater frequency than respondents who rated the same behaviours in the context of a 'work atmosphere' questionnaire. The authors concluded that the context in the first condition in which the survey was labeled as a sexual harassment survey had primed respondents to answer in a particular way. This use of context and inferring pragmatic meaning becomes more pronounced when respondents are unable to ask for clarification from the measure's administrator (Schwarz, 1999).

In the context of the present study, most participants herein identified the WAI as being representative of facets in therapy that they themselves believed to be important. Most participants across both phases, though particularly Phase One, described helpful elements of the alliance in line with those presented in the WAI (e.g., mutual respect, support, directed purpose, personal impact, etc.). It is unclear,

though, the degree to which participants would have expressed such views had they not been regularly completing this measure. Perhaps participants identified these facets as important due to beliefs that, if such elements were not helpful, the WAI would not be asking about them. A minority of participants in Phase One reported views that not all elements of the alliance queried by the WAI were important for them, such as a Phase One participant's reflection that her therapist's liking for her was inconsequential to treatment success.

The above reflection is important in light of the contention that self-report questionnaires can bias participants' responses by perpetuating researcher's own perspectives, thus influencing findings (Galsinski, 2008; Galasinski & Kozlowska, 2010). This is relevant in the present study where the majority of participants in Phase One identified 'helpful' elements in the relationship and treatment that were primarily in line with those queried in the WAI; these were described in the *Relating Together* main categories. It is unclear in this work whether a full sample of participants who were naïve to the WAI would provide responses consistent with the WAI. The literature regarding participants' perspectives on the alliance (Bachelor, 1995; Fitzpatrick et al., 2009; Hatchers & Barends, 1996) suggests that such naïve clients would offer similar reflections as to which factors were helpful, although the degree to which they would have done so in this sample is unknown. In essence, the question as to whether the WAI assesses what clients themselves believe to be important in therapy cannot be answered with confidence in this study.

In their investigation into respondents' experiences completing a questionnaire about loss of employment that utilized categorical response options, Galasinski and Kozlowska (2010) developed the understanding that their participants 'negotiated' tension between their experiences and the questionnaire utilizing one of three strategies: 1) Rejecting the questionnaire or its items outright (as being inapplicable or reductionist); 2) Matching their perspectives into the predetermined questionnaire categories; and 3) Reformulating the items to more accurately reflect respondents' experiences. Participants in the present work, particularly those providing broader reflections in Phase One interviews, described similar strategies to those identified by Galasinski and Kozlowska. Each of these points of convergence will be described in more detail below.

Related to Galasinski and Kozlowska's (2010) first strategy, several participants in both phases 'rejected' elements of the WAI as being inapplicable for them, such as the erroneous assumption identified in item 9 ("I feel _____ cares about me even when I do things that he/she does not approve of"), as well as the reduction of the complex therapeutic relationship to a quantitative scale identified in Phase One. In contrast with Galasinski and Kozlowska, no participant herein rejected the WAI outright or rejected the concepts of therapist and treatment effectiveness as being important areas for feedback.

Regarding their participants' second strategy, Galasinski and Kozlowska (2010) describe their understandings of respondents' experiences when attempting to 'match' respondents' lives to the measure under discussion. The paragraph from

their work is reproduced in its entirety below, as it strongly conveys the tension and negotiation at work when respondents completed the employment questionnaire:

> This tension can be resolved by constructing the experience in such a way that it can be described by the questionnaire's items. Thus, the respondents chose to either reconstruct their experience in a way that it can be represented as fitting into the categories offered or simply invalidate it by selecting any option…it is the pressure of the official task that must be completed. They feel obliged to do the task, even if it means reformulating stories of their experience. (Galasinski & Kozlowska, 2008, p. 279)

Their respondents worked to fit their experiences into the constraints of that questionnaire's items and assumptions. In this study, participants likewise described experiences whereby they needed to work to conform to the scale and its frame; however, most participants did not convey experiences that were notably critical of the reductionist nature of the scale and the need to dilute complex experiences into more simplistic terms to be amenable to quantitative rating.

In describing their respondents' third strategy, Galasinki and Kozlowska identified their participants as 'reformulating' the questionnaire items to both make them more intelligible or relevant for them as well as allowing for a more positive self-presentation to the researchers. The majority of participants in both phases described elements of this 'reformulation' process. A majority explicitly identified the tendency to reformulate some of the items or scale points of the WAI by making ambiguous words or items intelligible and meaningful for them, such as by substituting 'respects' for 'appreciates' to help answer item 7. A minority of Phase

One participants also described a tendency to provide a 'positive self-presentation' to their therapists or therapists' supervisors. A majority of Phase Two participants identified beliefs that some WAI items encouraged such a positive skew, though did not necessarily endorse doing so themselves.

Social Desirability and Alliance Ratings

This tendency towards 'positive self-presentation' highlighted in Galasinski and Kozlowska (2010) speaks to concerns about social desirability bias and its effect on survey responses, such as the WAI. The present work sheds some light on such questions about the role of a desirability bias at work in clients' WAI ratings, as encapsulated in the *I wanted to be honest* subcategory from Phase One. In a study investigating clients' ratings, Tryon and colleagues (2008) found that the ratings analyzed in their meta-analysis often represented only the upper 20% of the points on each rating scale (this work included the WAI and other alliance measures). Tryon et al. concluded that a social desirability bias might have led to inflation in the ratings and ceiling effects. The findings in the present study suggest, that while desirability bias might be a concern for some clients, the majority do not inflate their ratings in order to appear favourably to therapists or supervisors.

These results lend direct support to the conclusions made in another recent study investigating the hypothesized relationship between desirability and the alliance. Reese and colleagues (2013) tested this connection by analyzing desirability as related to the alliance on the WAI-SR and the Session Rating Scale (SRS; Miller, Duncan, & Johnson, 2000; as cited in Reese et al., 2013); the latter is a

four-item scale based in part on Bordin's (1979) tripartite alliance theory.[8] Reese et al. found that clients did not vary their ratings whether the therapist was present in the room during ratings, whether clients knew therapists would view the ratings, or whether therapists provided explicit feedback based on the ratings. These authors also found no significant correlation between ratings and a measure of social desirability (i.e., a measure assessing the degree to which clients wanted to appear favourably to others in general).

The present work offers possible answers to questions about the reasons behind the high alliance ratings encountered by Tryon and colleagues (2008). While some respondents acknowledged quite explicitly that they intentionally increased their ratings so as to allow them or their therapists to appear more positive than they might be otherwise, the majority of participants in both phases of this investigation denied doing this. The participants who denied inflating ratings reported that they thought that providing honest feedback to therapists and supervisors was more helpful, both for the clients to receive the most effective care and for the (mostly) trainee therapists to acquire the best training support. Virtually all of these participants also indicated that they had genuinely positive experiences with therapists and in treatment and that their higher ratings were accurate reflections of their attitudes.

On the face of it, participants' responses suggest that socially desirable responding did not influence their WAI ratings in this study. However, there are

---

[8] Three of the four items on the SRS correspond to each of Bordin's (1979) three factors. The fourth item relates to Hatcher and Barend's (1996) Confident Collaboration factor. Reese et al. (2013) found that the SRS and WAI-SR correlated moderately (r = .57-.65).

several possible ways to interpret this finding: 1) Only a minority of participants in this sample (two from Phase One) did in fact inflate their ratings, and the majority were accurate in their reports to the interviewer that they rated honestly and accurately; 2) A higher proportion of participants did inflate their ratings without reporting this information in interviews, but they inflated ratings sub/unconsciously and were thus unable to reflect on it explicitly; or 3) A higher proportion of participants did inflate ratings and did so consciously, but this desirability bias was also evidenced during interviews, and thus these participants did not disclose their rating inflation on the assumption that doing so was "undesirable" to the interviewer. It will be important for future studies to explore this issue more explicitly.

Clients' Pretreatment Expectations

Finally, the question as to how early session ratings of the alliance relate to clients' pretreatment expectations for therapists and therapy has been a key focus of the psychotherapy research literature. Several studies have identified correlations between pretreatment expectations for change and measures of the alliance (Constantino, Arnow, Blasey, & Agras, 2005; Joyce & Piper, 1998; Patterson et al., 2008; Patterson, Anderson, & Wei, 2014; Tsai, Ogrodniczuk, Sochting, & Mirmiran, 2014). Using the WAI-SR, Patterson and colleagues (2008) found that clients' self-rated pretreatment 'Personal Commitment' to the therapeutic process was predictive of each of the WAI-SR subscale ratings after the third session. They argued that clients prepared to engage in the work of therapy "are better able to establish a strong, collaborative, and productive relationship with the therapist"

(Patterson et al., 2008, p. 532). These authors found, however, that clients' pretreatment expectations regarding the 'Facilitative Conditions' in therapy, which included clients' expectations as to whether therapists would like them, be genuine and nurturing, and whether they would identify discrepancies in clients' thoughts and behaviours, was unrelated with all WAI-SR subscale scores at the third session. These authors were understandably surprised about the lack of association between Facilitative Conditions estimates and clients' Bond ratings. Following up on this work, Patterson et al. (2014) conducted mediation analyses, hypothesizing that the alliance was a mediator between clients' pretreatment expectations and therapeutic outcome. In this analysis, Patterson and colleagues (2014) found support for the alliance as mediator of pretreatment expectations and outcome, as well as establishing significant relationships among Personal Commitment to therapy and Facilitative Conditions factors as well as total WAI-SR ratings at the third session. The level of clients' pretreatment symptom severity was controlled in the mediation analyses, perhaps explaining the newfound association between Facilitative Conditions and WAI-SR total scores in their study.

The current study may offer some limited insight into this connection between clients' pretreatment expectations and subsequent alliance ratings. While these reflections were not captured as formal categories or codes within this study's analyses proper, several participants explicated the links between early session ratings and their expectations for their therapists and treatment. One participant, WAI_25, noted that she was not particularly hopeful in beginning treatment ("I wasn't thinking that this was going to help me change"), which led to her providing

lower early WAI ratings for change/goals-focused items. Another participant, WAI_23, who was described in the single case analysis, identified that her rating on item 1 ("As a result of these sessions I am clearer as to how I might be able to change") after the first session was based on her belief that treatment could be helpful ("I felt like there was potential for the future that I would feel good"). A third participant, WAI_19, shared a similar perspective regarding goals-focused WAI items, noting that her ratings on goals items after early sessions were based on her assumptions that her therapist would ultimately help her to identify and target goals on which to work ("That was the point of therapy so I assumed [my therapist] would do that"). Each of these reflections suggests that these participants had an awareness that their early ratings, at least for goals-related questions, were related to their expectations about the purpose of therapy and their hopes that personal change would result from treatment. Such findings could thus mean that the construct on which clients are basing their ratings may change depending on the stage of therapy in which they are rating (i.e., ratings given earlier in therapy may be based more strongly on expectancies, whereas ratings provided later in treatment may be more strongly rooted in actual treatment experiences).

Clinical Implications

The present study highlights the need for therapists and trainee therapists' supervisors to continue to have or begin to encourage discussions with clients about their experiences completing process measures. It seems likely that trainee therapists are already aware of the importance of such conversations, as Overington, Fitzpatrick, Hunsley and Drapeau (2015) found that trainees believe the second

most important use for progress measures is in helping facilitate discussions with clients (the first most important use was rated as tracking therapeutic change itself). Such discussions may lead to several useful outcomes. In particular, one helpful outcome would be an increased understanding of clients' ratings and perspectives on the directions being taken in therapy. This kind of conversation could be connected with general discussions of goals and progress, or lack thereof, within treatment. In many courses of therapy, these conversations will (hopefully) already be taking place, and thus the intention here would simply be to bring the topic of ratings and the rating process into the fold of these process conversations. A possible, albeit unsubstantiated, concern with having passive feedback mechanisms like rating scales is that therapists believe that such scales obviate the need for direct and open process conversations. It is important to note that rating scales are complements to effective therapeutic process conversations, not replacements for such communication. It is clear from the present work that rating scales do not present the full picture of clients' experiences.

A second possible benefit of therapist-directed conversations about clients' ratings would be allowing the therapist to address any misunderstandings about particular items or the use of the ratings within the therapeutic or research context. This could help to reduce some clients confusion with item terms or meaning and allow clients to provide more accurate ratings. Whether the ratings are being used for clinical or research purposes, the accuracy of the ratings as a feedback tool is predicated on the accuracy of clients' understandings of the scale. The present work demonstrates that clients' regularly have difficulties with understanding the

meaning of certain items and direct conversations could serve to minimize these. Whether enhancing clients' comprehension of these items will lead to changes in their actual ratings, however, is unknown at this time.

A further benefit to increasing therapist communication about these measures will be to raise awareness of the potential for client discomfort with the rating process. While participants in the present study reported that they generally were not made significantly uncomfortable by providing ratings, a majority expressed mild discomfort and one participant described significant discomfort when providing 'honest' ratings for specific items. These participant descriptions underscore that the rating procedure is still a complex clinical interaction that holds the possibility of adding to clients' distress. As with other interactions, understanding our clients' perspectives on measures could provide useful information regarding conceptualization of clients' difficulties, as well as encouraging therapists to help reduce any distress associated with completing ratings.

A further clinical implication of this study pertains to the need to provide clients who complete questionnaires with an introduction to each questionnaire, its items, and the ways in which it is and is not used. Most participants in this study reported that they had not received detailed introductions to each of the measures and their purposes.[9] The ambiguity of the use of the WAI ratings caused some of that

[9] It is possible of course that participants were initially more oriented to the WAI and its use than they remembered and/or conveyed in interview. This is not an entirely unlikely possibility given the delay of at least several months between most participants' first completion of the WAI and the research interview; however, the need to introduce these measures remains nonetheless.

client discomfort and could potentially be avoided or at least minimized through clear explanations about the measure and its purposes, particularly in training clinics such as those sampled in this study. One example of such clarification could be to orient clients to the fact that their therapists' performance in training courses was not contingent on results from client feedback measures per se.

Considerations for the WAI-SR

While it is not the intention of this work to suggest revisions to such a well-validated measure as the WAI-SR, some considerations do warrant highlighting as possible additions to the scale. A first possibility based on participants' feedback from this study involves adding a response option to the scale for 'Not Applicable.' Considering a number of participants noted that certain items were not relevant for them, either because they perceived it to be too early in the process to make a judgment or because the relevant situation had not arisen, adding an option like this could be of benefit. It would reduce the need for participants to leave inapplicable questions either blank or pressure them to provide a rating even if it is not an accurate one.

Another consideration would be to include explicit clarification about the intended reference period being queried by the WAI-SR. Participants reported being unclear as to whether the ratings were to be based on only the most recent session, previous sessions, or some combination thereof. Such clarification would be particularly beneficial for research purposes. It would produce greater consistency in comprehension across participants and allow for more confident aggregation of data.

Participants' feedback regarding confusion or ambiguity about some of the items on the WAI-SR, particularly items 3, 7, and 9 on the Bond subscale, raise concerns about the interpretability of ratings derived from those items. When the items are interpreted differently across respondents and when that diverging data becomes aggregated, the assumptions underlying that compilation are violated (Galasinski & Kozlowska, 2010; Rosenbaum & Valsiner, 2011). This presents a dilemma for a scale as well established as the WAI-SR since these findings would suggest that at least a few of the items could benefit from rephrasing and the measure be subjected to revalidation. How or even whether rephrasing of these items would impact the quantitative findings is unknown. What is known is that the confidence in the methodological underpinnings of the measure and the interpretations that follow could be strengthened in this way.

The observation from some participants that certain items were better assessed using a dichotomous scale bears consideration. As Mallinckrodt and Tekie (2016) identify, quantitative practices that utilize the ostensibly ordinal frequency data of the WAI for undertaking statistical analyses designed for interval data raises concerns on its own. The present findings extend this concern by indicating that at least some clients treat some of the items as dichotomous, which may present further statistical challenges. Related to the points on the rating scale, Mallinckrodt and Tekie's suggestion of removing the ordinal frequency descriptors entirely from the Likert scale and only keeping a 5-point numerical scale is an interesting one. This proposal is supported by some of this study's participants who described difficulties distinguishing among points on the rating scale (e.g., 'Sometimes' versus

'Seldom'). It is unclear whether their suggestion of removing the frequency descriptions would help to avoid the issue of dichotomous ratings noted here.

Limitations of the Present Study

There are several limitations evident in the present study. A first concerns the composition of the samples. All participants had completed numerous therapy sessions prior to their interviews. These clients also had relatively strong alliances with their therapists. As such, the participants are not necessarily representative of other types of clients found in different settings. Completing interviews with clients who had already dropped out of treatment or who had sustained treatment but evinced less positive alliances could have impacted the results. As one Phase Two participant noted, if he did not think his therapist liked him, "I wouldn't say that [on the WAI], I just wouldn't show up the next time." As all participants were adults and no participants were mandated to attend treatment, the samples were thus comprised of clients who had found therapy meaningful in some way. Presumably, those clients who continued to attend sessions and participate in these interviews had established connections with their therapists and courses of treatment, which likely would have differed from clients who terminated treatment earlier.

A related note concerning the sample pertains to the participants' experiences with the WAI. All but one participant had regularly completed WAIs (usually after every second session). Twenty-one of 22 participants were thus seasoned WAI completers. Undertaking these interviews with a full sample of clients inexperienced with the WAI might have yielded quite different results and should be explored in future studies.

Another limitation concerns participant and therapist characteristics and the contexts from which participants were recruited. The majority of participants were undergraduate students. Additionally, all but one therapist was under supervision at the time of interviews, and the majority were novice practitioners meeting with clients as a part of graduate practicum courses. Participants were aware of therapists' training status, which may have impacted the ways in which they completed the measure or what they were comfortable disclosing in interview. For some participants, their belief that therapists were being graded on WAI ratings put some degree of pressure on their ratings, which may have altered the accuracy with which they completed WAIs or described doing so.

During Phase Two analyses, only a small proportion of interview data was captured by the predefined CCS coding system; thus Phase Two interviews may have demonstrated a higher degree of convergence or divergence in findings when compared to Phase One had all interview data been utilized. This was beyond the scope of the present study and would have required a different analytic procedure, such as Grounded Theory or thematic analysis. Capturing all of Phase Two interview accounts may have modified results in some unknown fashion. While this was beyond the scope of the present study, a reanalysis of these interviews could be considered for a future work.

Directions for Future Research

There are a number of avenues for future research that are suggested by the present work. One area of interest that would help to enhance our understanding of clients' rating processes would involve undertaking something

akin to Interpersonal Process Recall interviews with participants following sessions. This approach could entail having participants complete the WAI and identify specific, salient events in that most recent therapy session that were the basis for their ratings. This design would provide a deeper understanding of how the 'critical incidents' of alliance formation that clients experience in therapy (Fitzpatrick et al., 2006) influence their ratings of that alliance on quantitative measures.

Another direction for future research pertains to participant selection. The samples in the present work were comprised primarily of longer-term therapy clients who had completed the WAI on a number of occasions. Considering the finding in this work that repeated completions of the measure induce clients to provide ratings more quickly, with more predictability, and with less explicit thinking about each item, conducting interviews with clients who are in the early-phase of therapy and/or have never completed a WAI previously could yield quite different results.

An additional direction pertains to participants in both phases identifying confusion about the time period being queried by the WAI. Considering participants in Phase One described different interpretations of this confusion and subsequent approaches to provide ratings (i.e., by only rating based on the recent session, only rating on the totality of treatment, or rating based on some combination of the two), future studies could examine how varying explicit instructions regarding the time period impacts the ratings provided. For example, a study could involve two groups of respondents with the first being asked to rate only on the most recent session and the second being asked to reflect on treatment as a whole.

This study's uncovering of the complexities evident throughout the WAI-SR rating process highlights the importance of applying similar methods to other psychotherapy process and progress measures. These methods would be applicable during both the development phase of new measures and in revisiting more established scales, as in this work, to gain a better understanding of what is being queried and how respondents perceive each scale. In light of recommendations for the broader application of progress measures in clinical settings (Overington & Ionita, 2012), investigation into clients' qualitative perceptions of these measures would be of considerable clinical utility.

Conclusions

The present study raises important considerations for the psychotherapy clinical and research literatures. The results highlight how participants identify the key purpose of the scale as helping to measure and communicate the degree to which they believe that they and their therapists are working together effectively towards common therapeutic goals. In line with perspectives put forward by Stiles and Goldsmith (2010), participants viewed the WAI as a means with which to evaluate their therapists and treatments in general terms. In essence, participants seemed to be answering two broad questions when providing their ratings: 1) Are my therapist and I relating in a comfortable and respectful manner?; and 2) Are we working together effectively towards my goals? Answers to the former question appear to be communicated via clients' Bond ratings, while answers to the latter are more specifically addressed through clients' Goals and Tasks ratings. These two general questions can be further distilled into a single question that participants

appear to answer with their ratings: Are my therapist and I on the right track in treatment? The total score of their ratings on the scale conveys the answers to this question.

This study also contributes to the field by offering the first perspective into clients' qualitative processes of engagement with the widely used WAI-SR. Participants in this study identify the WAI-SR to be a generally intelligible and efficient means with which to collect information about therapist and treatment effectiveness, but that it is also a measure with several elements that lead to confusion for those who complete it, notably related to item ambiguity and an unclear reference period queried by the scale. These ambiguities in clients' understandings of certain items and the relevant reference period call into question how confident researchers can be in the accuracy of the ratings that clients provide. The existence of these potential sources of error for this widely used psychotherapy instrument has not been demonstrated before in the literature.

Considering the prominence of the alliance construct in the literature and the prevalence with which the WAI and its successors are utilized, there are numerous implications for psychotherapy research and practice that follow from the present work. The findings introduce evidence for several sources of error in clients' WAI ratings, which serves to call the accuracy of these ratings into question. These findings also cast doubt on the influence of social desirability concerns on such ratings and elucidate the need for the recognition that clients' completion of psychotherapy measures remains a part of the clinical encounter. Perhaps most importantly, the study underscores the need to employ qualitative interview and

analytic methods when developing and/or utilizing quantitative scales to ensure that respondents understand the measures in the manner intended.

References

Adair, C.E., Holland, A.C., Patterson, M.L., Mason, K.S., Goering, P.N., Hwang, S.W., & At Home/Chez Soi Project Team. (2011). Cognitive interviewing methods for questionnaire pre-testing in homeless persons with mental disorders. *Journal of Urban Health: Bulletin of the New York Academy of Medicine, 89*(1), 36-52.

Alexander, L.B., & Luborsky, L. (1987). The Penn Helping Alliance Scales. In L.S. Greenberg & W.M. Pinsoff (Eds.), *The psychotherapeutic process: A research handbook* (pp. 325-366). New York: Guilford Press.

Bachelor, A. (1995) Clients' perception of the therapeutic alliance: A qualitative analysis. *Journal of Counseling Psychology, 42*(3), 323-337.

Bachelor, A. (2013). Clients' and therapists' views of the therapeutic alliance: Similarities, differences and relationship to therapy outcome. *Clinical Psychology and Psycotherapy, 20,* 118-135.

Bachelor, A., Meunier, G., Laverdiere, O., & Gamache, D. (2010). Client attachment to therapist: Relation to client personality and symptomatology, and their contributions to the therapeutic alliance. *Psychotherapy Theory, Research, Practice, Training, 47*(4), 454-468.

Barroso, J., & Sandelowski, M. (2001). In the field with the Beck Depression Inventory. *Qualitative Health Research, 11*(4), 491-504.

Beatty, P.C., & Willis, G.B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly, 71*(2), 287-311.

Bedi, R.P., Davis, M.D., & Williams, M. (2005). Critical incidents in the formation of the therapeutic alliance from the client's perspective. *Psychotherapy: Theory, Research, Practice, Training, 42,* 311-323.

Beretta, V., de Roten, Y., Stigler, M., Drapeau, M., Fischer, M., & Despland, J. (2005). The influence of patient's interpersonal schemas on early alliance building. *Swiss Journal of Psychology, 64*, 13-20.

Bocklisch, F., Bocklisch, F.S., & Krems, J. (2012). Sometimes, often, and always: Exploring the vague meanings of frequency expressions. *Behavior Research & Methods, 44*(1), 144-157.

Bordin, E.S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy Theory, Research, and Practice, 16*(3), 252-260.

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research, 12,* 229-238.

Constantino, M.J., Arnow, B.A., Blasey, C., & Agras, W.S. (2005). The association between patient characteristics and the therapeutic alliance in cognitive-behavioral and interpersonal therapy for bulimia nervosa. *Journal of Consulting and Clinical Psychology, 73,* 203-211.

Corbin, J., & Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (3rd ed.). Thousand Oaks, CA: Sage.

Denzin, N.K. (2012). Triangulation 2.0. *Journal of Mixed Methods Research, 6*(2), 80-88.

Ellingson, L.L. (2009). *Engaging crystallization in qualitative research: An introduction.* Thousand Oaks, CA: Sage.

Ellingson, L.L. (2014). "The truth must dazzle gradually": Enriching relationship research using a crystallization framework. *Journal of Social and Personal Relationships, 31*(4), 442-450.

Elvins, R., & Green, J. (2008). The conceptualization and measurement of the therapeutic alliance: An empirical review. *Clinical Psychology Review, 28,* 1167-1187.

Falkenstrom, F., Granstrom, F., & Holmqvist, R. (2014). Working alliance predicts psychotherapy outcome even while controlling for prior symptom improvement. *Psychotherapy Research, 24,* 146-159.

Farmer, T., Robinson, K., Elliott, S.J., & Eyles, J. (2006). Developing and implementing a triangulation protocol for qualitative health research. *Qualitative Health Research, 16*(3), 377-394.

Fassinger, R.E. (2005). Paradigms, praxis, problems, and promise: Grounded theory in counseling psychology research. *Journal of Counseling Psychology, 52*(2), 156-166.

Ferenczi, S. (1988). The clinical diary of Sandor Ferenczi. In J. Dupont (Ed.). Cambridge, MA: Harvard University Press. (Original work published 1932)

Fitzpatrick, M.R., Iwakabe, S., & Stalikas, A. (2005). Perspective divergence in the working alliance. *Psychotherapy Research, 15*(1-2), 69-79.

Fitzpatrick, M.R., Janzen, J., Chamodraka, M., & Park, J. (2006). Client critical incidents in the process of early alliance development: A positive emotion-exploration spiral. *Psychotherapy Research, 16,* 486-498.

Fitzpatrick, M.R., Janzen, J., Chamodraka, M., Gamberg, S., & Blake, E. (2009). Client relationship incidents in early therapy: Doorways to collaborative engagement. *Psychotherapy Research, 19*(6), 654-665.

Freud, S. (1990). The dynamics of transference. In A. H. Esman (Ed.), *Essential papers on transference.* New York: New York University Press. (Original work published 1912)

Galasinski, D. (2008). Constructions of the self in interactions with the Beck Depression Inventory. *health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine, 12*(4), 515-533.

Galasinski, D. & Kozlowska, O. (2010). Questionnaires and lived experiences: Strategies of coping with the quantitative frame. *Qualitative Inquiry, 16*(4), 271-284.

Galesic, M., & Tourangeau, R. (2007). What is sexual harassment? It depends on who asks! Framing effects on survey responses. *Applied Cognitive Psychology, 21*, 189-202.

Gaston, L. (1991). Reliability and criterion-related validity of the California Psychotherapy Alliance Scales-Patient Version. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*(1), 68-74.

Gelso, C.J., Perez-Rojas, A.E., & Marmarosh, C. (2014). Love and sexuality in the therapeutic relationship. *Journal of Clinical Psychology: In Session, 70*(2), 123-134.

Gilgun, J.F. (2004). Qualitative methods and the development of clinical assessment tools. *Qualitative Health Research, 14*(7), 1008-1019.

Glaser, B.G., & Straus, A. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago, IL: Aldine.

Grant, M.A., Rohr, L.N., & Grant, J.T. (2012). How informants answer questions: Implications for reflexivity. *Field Methods, 24*(2), 230-246.

Greenson, R. (1990). The working alliance and the transference neurosis. In A. H. Esman (Ed.), *Essential papers on transference.* New York: New York University Press. (Original work published 1965)

Greenson, R.R. (1967). *The Technique and Practice of Psychoanalysis* (Vol. 1). New York, NY: International Universities Press.

Greenson, R. (1971). The real relationship between the patient and the psychoanalyst. In M. Kanzer (Ed.), *The unconscious today* (pp. 213-232). New York: International Universities Press.

Groleau, D., Young, A., & Kirmayer, L.J. (2006). The McGill Illness Narrative Interview (MINI): An interview schedule to elicit meanings and modes of reasoning related to illness experience. *Transcultural Psychiatry, 43*(4), 671-691.

Hatcher, R.L. (2010). Alliance theory and measurement. In J.C. Muran & J.P Barber (Eds.), The therapeutic alliance: An evidence-based guide to practice (pp. 7-28). New York: Guildford Press.

Hatcher, R.L., & Barends, A.W. (1996). Patients' view of the alliance in psychotherapy: Exploratory factor analysis of three alliance measures. *Journal of Consulting and Clinical Psychology, 64*(6), 1326-1336.

Hatcher, R.L., & Gillaspy, J.A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research, 16*(1), 12-25.

Hatcher, R.L., Barends, A.W., Hansell, J., & Gutfreund, M.J. (1995). Patients' and therapists' shared and unique views of the therapeutic alliance: An investigation using confirmatory factor analysis in a nested design. *Journal of Consulting and Clinical Psychology, 63*(4), 636-643.

Horvath, A.O. (1981). *An exploratory study of the working alliance: Its measurement and relationship to therapy outcome* (Unpublished doctoral dissertation). University of British Columbia, Vancouver.

Horvath, A.O. (2001). The alliance. *Psychotherapy: Theory, Research, Practice, Training, 38,* 365-372.

Horvath, A.O., Del Re, A.C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48*(1), 9-16.

Horvath, A. & Bedi, R. (2002). The alliance. In J.C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp.37-69). London: Oxford University Press.

Horvath, A., & Greenberg, L.S. (1989). Development and validation of the working alliance inventory. *Journal of Counseling Psychology, 36,* 223-233.

Horvath, A.O. & Symonds, B.D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38,* 139-149.

Iacoviello, B.M., McCarthy, K.S., Barrett, M.S., Rynn, M., Gallop, R. & Barber, J.P. (2007). Treatment preferences affect the therapeutic alliance: Implications for randomized controlled trials. *Journal of Consulting and Clinical Psychology, 75,* 194-198.

Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12-19.

Joyce, A.S., & Piper, W.E. (1998). Expectancy, the therapeutic alliance, and treatment outcome in short-term individual psychotherapy. *The Journal of Psychotherapy Practice and Research, 7,* 236-248.

Joyce, A.S., Ogrodniczuk, J.S., Piper, W.E., & McCallum, M. (2003). The alliance as mediator of expectancy effects in short-term individual therapy. *Journal of Consulting and Clinical Psychology, 71*(4), 672-679.

Kagan, F. (2014). *Untying the knot: A qualitative inquiry into the enduring marriage* (Unpublished doctoral dissertation). York University, Toronto.

Karabenick, S.A., Woolley, M.E., Friedel, J.M., Ammon, B.V., Blazevski, J., Bonney, C.R.,…Musu, L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist, 42*(3), 139-151.

Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing.* Thousand Oaks, CA: Sage.

Lincoln, Y., & Guba, E. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N. Denzin & Y. Lincoln (Eds.), *Handbook of Qualitative Research, Second Edition* (pp. 163-188). Thousand Oaks, CA: Sage.

MacFarlane, P., Anderson, T., & McClintock, A.S. (2015). The early formation of the working alliance from the client's perspective: A qualitative study. *Psychotherapy, 52*(3), 363-372.

Mallinckrodt, B., & Tekie, Y.T. (2016). Item response theory of Working Alliance Inventory, revised response format, and new Brief Alliance Inventory. *Psychotherapy Research, 26*(6), 694-718.

Marmar, C. R., Weiss, D.S., & Gaston, L. (1989). Toward the validation of the California Therapeutic Alliance Rating System. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*(1), 46-52.

Martin, D.J., Garske, J.P. & Davis, M.K. (2000). Relation of therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 68,* 438-450.

McHugh, M.L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276-282.

Meijer, P.C., Verloop, N., & Beijaaard, D. (2002). Multi-method triangulation in a qualitative study on teachers' practical knowledge: An attempt to increase internal validity. *Quality & Quantity, 36,* 145-167.

Miles, M.B., & Huberman, A.M. (1994). *Qualitative Data Analysis*. Thousand Oaks, CA: Sage.

Miller, S.D., Duncan, B.L., & Johnson, L.D. (2000). *The Session Rating Scale 3.0*. Chicago, IL: Authors.

Muran, J.C., & Barber, J.P. (2010). An introduction. In J.C. Muran & J.P Barber (Eds.), The therapeutic alliance: An evidence-based guide to practice (pp. 1-4). New York: Guildford Press.

O'Malley, S.S., Suh, C.S., & Strupp, H.H. (1983). The Vanderbilt Psychotherapy Process Scale: A report on the scale development and a process-outcome study. *Journal of Consulting and Clinical Psychology, 51*(4), 581-586.

Overington, L., & Ionita, G. (2012). Progress monitoring measures: A brief guide. *Canadian Psychology, 53*(2), 82-92.

Overington, L., Fitzpatrick, M., Hunsley, J., & Drapeau, M. (2015). Trainees' experiences using progress monitoring measures. *Training and Education in Professional Psychology, 9*(3), 202-209.

Patterson, C.L., Anderson, T., & Wei, C. (2014). Clients' pretreatment role expectations, the therapeutic alliance, and clinical outcomes in outpatient therapy. *Journal of Clinical Psychology, 70*(7), 673-680.

Patterson, C.L., Uhlin, B., & Anderson, T. (2008). Clients' pretreatment counseling expectations as predictors of working alliance. *Journal of Counseling Psychology, 55*(4), 528-534.

Ponterotto, J.G. (2005). Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science. *Journal of Counseling Psychology, 52*(2), 126-136.

Priede, C., & Farrall, S. (2011). Comparing results from different styles of cognitive interviewing: 'Verbal probing' vs. 'thinking aloud.' *International Journal of Social Research Methodology, 14*(4), 271-287.

Prochaska, J.O. & Norcross, J.C. (2002). Stages of change. In J.C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 303-313). London: Oxford University Press.

Reese, R.J., Gillaspy, J.A., Owen, J.J., Flora, K.L., Cunningham, L.C., Archie, D., & Marsden, T. (2013). The influence of demand characteristics and social desirability on clients' ratings of the therapeutic alliance. *Journal of Clinical Psychology, 69*(7), 696-709.

Rennie, D.L. (2006). The grounded theory method: Application of a variant of its procedure of constant comparative analysis to psychotherapy research. In C.T. Fischer (Ed.), *Qualitative Research Methods for Psychologists: Introduction Through Empirical Studies* (pp. 59-78). Boston, MA: Academic Press.

Rosenbaum, P.J., & Valsiner, J. (2011). The un-making of a method: From rating scales to the study of psychological processes. *Theory & Psychology, 21*(1), 47-65.

Rothgeb, J., Willis, G., & Forsyth, B. (2005). Questionnaire pretesting methods: Do different techniques and different organizations produce similar results? Washington, DC:

U.S. Census Bureau. Retrieved from https://www.census.gov/srd/papers/pdf/rsm2005-02.pdf.

Safran, J.D. & Muran, J.C. (2000). *Negotiating the therapeutic alliance: A relational treatment guide.* New York: Guilford Press.

Safran, J.D. & Muran, J.C. (2006). Has the concept of the therapeutic alliance outlived its usefulness? *Psychotherapy: Theory, Research, Practice, Training, 43*(3), 286-291.

Samstag, L.W., Muran, J.C., Wachtel, P.L., Slade, A., Safran, J.D., & Winston, A. (2008). Evaluating negative process: A comparison of working alliance, interpersonal behavior, and narrative coherency among three psychotherapy outcome conditions. *American Journal of Psychotherapy, 62*(2), 165-194.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*(2), 93-105.

Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology, 21,* 277-287.

Smaling, A. (1987). Methodologische objectiviteit en kwalitatief onderzoek [Methodological objectivity and qualitative research]. Lisse, The Netherlands: Swets & Zeitlinger.

Smith, M.V. (2008). Pain experience and the imagined researcher. *Sociology of Health & Illness, 30*(7), 992-1006.

Sterba, R. (1934). The fate of the ego in analytic therapy. *International Journal of Psychoanalysis, 15,* 117-125.

Sterba, R. (1990). The dynamics of the dissolution of the transference resistance. In A. H. Esman (Ed.), *Essential papers on transference.* New York: New York University Press. (Original work published 1940)

Stiles, W.B., & Goldsmith, J.Z. (2010). The alliance over time. In J.C. Muran & J.P Barber (Eds.), The therapeutic alliance: An evidence-based guide to practice (pp. 44-62). New York: Guildford Press.

Tourangeau, R., & Rasinski, K.A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103*(3), 299-314.

Tourangeau, R., Ripps, L.J., & Rasinski, K. (2000). *The Psychology of Survey Response.* Cambridge, UK: Cambridge University Press.

Tracey, T.J. & Kokotovic, A.M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1,* 207-210.

Tryon, G.S., Blackwell, S.C., & Hammel, E.F. (2007). A meta-analytic examination of client-therapist perspectives of the working alliance. *Psychotherapy Research, 17*(6)*,* 629-642.

Tryon, G.S., Blackwell, S.C., & Hammel, E.F. (2008). The magnitude of client and therapist working alliance ratings. *Psychotherapy: Theory, Research, Practice, Training, 45*(4), 546-551.

Tsai, M., Ogrodniczuk, J.S., Sochting, I., & Mirmiran, J. (2014). Forecasting success: Patient's expectations for improvement and their relations to baseline, process, and outcome variables in group cognitive-behavioral therapy for depression. *Clinical Psychology and Psychotherapy, 21*, 97-107.

Turner, H., Bryant-Waugh, R., & Marshall, E. (2015). The impact of early symptom change and therapeutic alliance on treatment outcome in cognitive-behavioural therapy for eating disorders. *Behaviour Research and Therapy, 73,* 165-169.

Westra, H.A., Constantino, M.J., & Antony, M.M. (2016, March 17). Integrating motivational interviewing with cognitive-behavioral therapy for severe Generalized Anxiety Disorder: An allegiance-controlled randomized clinical trial. *Journal of Consulting and Clinical Psychology.* Advance online publication. http://dx.doi.org/10.1037/ccp0000098

Willis, G.B. (2004). Cognitive interviewing revisited: A useful technique, in theory? In S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 23-44). Hoboken, NJ: John Wiley & Sons.

Xu, H., & Tracey, T.J.G. (2015). Reciprocal influence model of working alliance and therapeutic outcome over individual therapy course. *Journal of Counseling Psychology, 62*(3), 351-359.

Zetzel, E.R. (1990). Current concepts of transference. In A. H. Esman (Ed.), *Essential papers on transference.* New York: New York University Press. (Original work published 1956)

Appendices

Appendix A: Working Alliance Inventory - Short Revised

Instructions: Below is a list of statements and questions about experiences people might have with their therapy or therapist. Some items refer directly to your therapist with an underlined space – as you read the sentences, mentally insert the name of your therapist in place of _____ in the text. Think about your experience in therapy, and decide which category best describes your own experience.

IMPORTANT!!! Please take your time to consider each question carefully.

1) "As a result of these sessions I am clearer as to how I might be able to change"

1	2	3	4	5
Seldom	Sometimes	Fairly Often	Very Often	Always

2) "What I am doing in therapy gives me new ways of looking at my problem"

5	4	3	2	1
Always	Very Often	Fairly Often	Sometimes	Seldom

3) "I believe _____ likes me"

1	2	3	4	5
Seldom	Sometimes	Fairly Often	Very Often	Always

4) "_____ and I collaborate on setting goals for my therapy"

1	2	3	4	5
Seldom	Sometimes	Fairly Often	Very Often	Always

5) "_____ and I respect each other"

5	4	3	2	1
Always	Very Often	Fairly Often	Sometimes	Seldom

6) "_____ and I are working towards mutually agreed upon goals"

5	4	3	2	1
Always	Very Often	Fairly Often	Sometimes	Seldom

7) “I feel that _____ appreciates me”

1	2	3	4	5
Seldom	Sometimes	Fairly Often	Very Often	Always

8) “_____ and I agree on what is important for me to work on”

5	4	3	2	1
Always	Very Often	Fairly Often	Sometimes	Seldom

9) “I feel that _____ cares about me even when I do things he/she does not approve of”

1	2	3	4	5
Seldom	Sometimes	Fairly Often	Very Often	Always

10) “I feel the things that I do in therapy will help me to accomplish the changes that I want”

5	4	3	2	1
Always	Very Often	Fairly Often	Sometimes	Seldom

11) “_____ and I have established a good understanding of the kind of changes that would be good for me”

5	4	3	2	1
Always	Very Often	Fairly Often	Sometimes	Seldom

12) “I believe the way in which we are working on my problem is correct”

1	2	3	4	5
Seldom	Sometimes	Fairly Often	Very Often	Always

Appendix B: Participant Informed Consent Form

Clients' Perspectives of the Working Alliance Inventory
Informed Consent Form

Study Name: Clients' Perspectives of the Working Alliance Inventory

Researcher: James Watson-Gaze, M.A.

What You Will Be Asked to Do in the Research: This study aims to understand clients' perspectives of the Working Alliance Inventory-Short Revised (WAI-SR), a post-session measure that assesses the working relationship between clients and therapists. You will be asked to take part in a single interview for which you will bring in your most recently completed WAI-SR. During the interview, you will be asked to discuss aspects of your awareness when you fill out specific items on the WAI-SR. Additionally, you will be asked general questions about your experience with the WAI-SR. A sample question for the interview is, "Overall, how well do you feel this measure captures your experience of your working relationship with your therapist?" Interviews will last approximately 30-60 minutes.

Risks and Discomforts: We do not foresee any risks or discomfort associated with your participation in the research.

Benefits of the Research and Benefits to You: Clients' feedback about their therapy experiences has proven to be very helpful for improving treatment outcomes. Your feedback in this interview will help us understand the WAI-SR and its ability to measure client perceptions of the working relationship between clients and therapists. Results of this study will be disseminated through traditional academic media (e.g., journal articles). As this project is unfunded, you will not receive financial remuneration for your participation.

Voluntary Participation: Your participation in the study is completely voluntary and you may choose to stop participating at any time. If you choose to withdraw, all data from your interview will be destroyed immediately. Your decision to not volunteer or to withdraw will not influence the treatment you are receiving or your relationship with your therapist or York University in any way, either now or in the future.

Withdrawal from the Study: You can stop participating in the interview at any time, for any reason, if you so decide. Your decision to stop participating, or to refuse to answer particular questions, will not affect your treatment or your relationship with your therapist or York University in any way.

Confidentiality: All information you supply during the research will be held in confidence and your name will not appear on the transcript of your interview or in any report or publication of the research. The audio recording of your interview will be destroyed once it has been transcribed. The transcript of your interview will be safely stored in a locked

facility for a period of five years and will be destroyed after this time. Only research team members will have access to this information. The confidentiality of your interview will be protected to the fullest extent of the law.

Questions About the Research? If you have questions about the research in general or about your role in the study, please feel free to contact either James Watson-Gaze or the Graduate Office, Department of Psychology, York University. This research has been reviewed by the Human Participants in Research Committee and York University's Ethics Review Board and conforms to the standards of the Canadian Tri-Council Research Ethics guidelines. If you have any questions about this process or about your rights as a participant in the study, please contact Ms. Alison Collins-Mrakas, Manager, Office of Research Ethics.

Legal Rights and Signatures:

I, _____ (please print), consent to participate in the "Clients' Perspectives of the Working Alliance Inventory" study conducted by James Watson-Gaze. I have understood the nature of this project, and I wish to participate. I am not waiving any of my legal rights by signing this form. My signature below indicates my consent.

☐ I give permission for the anonymized data from my interview to be used for future research purposes. I understand that any future research will ensure the confidentiality of my responses

Participant Signature _____ **Date** _____

Investigator Signature _____ **Date** _____

Appendix C: Demographic Questionnaire

Client I.D. _____

Age _____

Gender:

Male	_____
Female	_____
Transgender	_____

Highest educational level:

___ junior high
___ some high school
___ high school diploma or GED
___ technical school
___ some university/college
___ associate's degree
___ bachelor's degree
___ graduate degree

Marital status:

___ single, never married
___ single, never married, in a relationship
___ single, never married, living with a partner
___ married
___ separated
___ divorced
___ widowed
___ remarried

Are you a York student? Yes No

Do you pay a fee for your sessions? Yes No

How many sessions have you had in your current course of therapy? _____

Have you ever had psychotherapy before this? Yes No

If yes, for approximately how long? _____

Appendix D: Tables for Phase Two Results

Table D4

Frequency of CCS Codes

CCS Code Type	Frequency
COMPREHENSION AND COMMUNICATION	**67**
Interviewer Difficulties	**0**
Inaccurate instructions	0
Complicated instructions	0
Difficult for interviewer to administer	0
Question Content	**50**
Vague topic/unclear question	12
Complex topic	9
Topic carried over from previous question	1
Undefined/vague term	28
Question Structure	**13**
Transition needed	0
Unclear instruction to respondent	0
Question too long	0
Complex/awkward syntax	2
Erroneous assumption	9
Several questions	2
Reference Period	**4**
Reference period carried over	0
Undefined reference period	4
Unanchored/rolling reference period	0
RETRIEVAL FROM MEMORY	**0**
Shortage of memory cues	0
High detail required/information unavailable	0
Long recall/reference period	0
JUDGMENT AND EVALUATION	**14**
Complex estimation	0
Potentially sensitive/desirability bias	14
RESPONSE SELECTION	**13**
Response Terminology	**1**
Undefined term(s) in response category	0
Vague term(s) in response category	1
Response Units	**4**
Response categories use wrong or mismatched units	4
Unclear to respondent what response categories are	0
Response Structure	**8**
Overlapping response categories	0
Missing response categories	8
TOTAL	**94**

Table D5

Frequency of Total CCS Codes by WAI-SR Item

CCS Code	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	**TOTAL**
COMPREHENSION AND COMMUNICATION													**77**
Interviewer Difficulties													**0**
Inaccurate instructions													0
Complicated instructions													0
Difficult for interviewer to administer													0
Question Content													**53**
Vague topic/unclear question			5		1		6						12
Complex topic	1		5				2		1			2	11
Topic carried over from previous question											1	1	2
Undefined/vague term			2	2			14				2	8	28
Question Structure													**16**
Transition needed													0
Unclear instruction to respondent													0
Question too long													0
Complex/awkward syntax											2		2
Erroneous assumption				1		3			6			2	12
Several questions									2				2
Reference Period													**8**
Reference period carried over													0
Undefined reference period		1	1	1	1	1			1		1	1	8
Unanchored/rolling reference period													0
RETRIEVAL FROM MEMORY													**0**
Shortage of memory cues													0
High detail required/information unavailable													0
Long recall/reference period													0
JUDGMENT AND EVALUATION													**15**
Complex estimation													0
Potentially sensitive/desirability bias	2		4	2			2			1		4	15
RESPONSE SELECTION													**16**
Response Terminology													**1**
Undefined term(s) in response category													0
Vague term(s) in response category			1										1
Response Units													**4**
Response categories use wrong units	1				1		1				1		4
Unclear response categories													0
Response Structure													**11**
Overlapping response categories													0
Missing response categories				1	1	2			5	1		1	11
TOTAL	4	1	18	7	4	6	25	0	15	2	7	19	**108**

Table D6
Frequency of CCS Codes Generated by Each Participant by WAI-SR Item

Participant	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	**TOTAL**
WAI_23	2						6		1				**9**
WAI_24			7		2		3		2		1	1	**16**
WAI_25									2	1		3	**6**
WAI_26	2			2			1					2	**7**
WAI_27				3		6	7		5	1		6	**28**
WAI_28													**0**
WAI_29		1	5	2	2		4		2		1		**17**
WAI_31			2						2		5	5	**14**
WAI_32			2				2		1			2	**7**
WAI_33			2				2						**4**
TOTAL	4	1	18	7	4	6	25	0	15	2	7	19	**108**