THESIS

SESSION SPECIFIC MEASUREMENT OF THE WORKING ALLIANCE

Submitted by

Joshua Powner

Department of Human Development and Family Studies

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2020

Master's Committee:

Advisor: Kelley Quirk

Gloria Luong
Jennifer Harman

ABSTRACT

SESSION SPECIFIC MEASUREMENT OF THE WORKING ALLIANCE

The psychotherapy working alliance is a dynamic construct that may exhibit periods of strength or periods of strain. However, current methods to assess the working alliance is problematic in identifying fluctuations. Assessing fluctuations may allow for better tracking of the working alliance, which in turn may help therapists to tailor their approaches accordingly. The current study developed and evaluated the Session Specific Alliance Measure (SSAM), examining its reliability as well as construct, concurrent, and predictive validity. A small sample (N = 47) of clients treated by 14 therapists were surveyed at a university health clinic. Findings in the current study reflect only one wave of data, despite attempts to collect multiple waves of data. These findings support the SSAM to reliably assess the working alliance as well as support its construct validity. However, findings failed to support the SSAM's concurrent and predictive validity. Limitations and future directions are discussed.

Keywords: working alliance, psychotherapy, reliability, validity, session specific alliance measure (SSAM)

TABLE OF CONTENTS

Session Specific Measurement of the Working Alliance

The psychotherapy working alliance is a dynamic construct that may exhibit periods of strength or periods of strain (Falkenström, Ekeblad, & Holmqvist, 2016). These fluctuations in strength may be seen in any or all of the working alliance's three subcomponents, which Bordin (1979) defines as (1) agreement on the goals of the therapy, (2) the agreement on the tasks or means to achieve those goals, and finally (3) the bond—which is the degree to which a client experiences warmth, trust, and affection toward and with the therapist (see Figure 1). An example illustrative of a strong alliance in these subcomponents is a client who seeks out therapy for their anxiety, who describers their goal as increasing their understanding of their anxiety in order to reduce how intensely they experience it each day. A therapist would understand and agree with this goal, and then the client and therapist would agree on a task of recording in a journal the types of stimuli (or triggers) that make them feel the most anxious. As the therapist calmly listens to the client's anxieties without criticism, this may foster feelings of trust and warmth – indicative of a strong bond. However, despite current measures that assess the working alliance's strength, no scale currently exists to measure the fluctuations of that strength over time. It is the aim of this paper to develop and test the initial reliability and validity of a measure that could assess fluctuations in the working alliance.

Many measures of the working alliance have been developed and empirically supported (McLeod & Weisz, 2005; Duncan, 2012; Munder, Wilmers, Leonhart, Linster, & Barth, 2010; Horvath & Greenberg, 1989). Although there is variation in their approach, they each generally elicit responses unspecific to the working alliance in any given session- termed here as "global responses". One item taken from the more commonly used alliance measure is "*I feel ___ cares about me even when I do things that he/she does not approve of"* (Horvath & Greenberg, 1989;

Munder et al., 2010). This item and others like it may be problematic because clients may feel cared for from their therapists overall but may vary in how cared for they feel during a particular session. This highlights a need for working alliance measurement approaches that allow for clients to articulate perceived fluctuations in any of the alliance components.

Given this need, it is not surprising that therapists are often still not aware of strains on the working alliance (Eubanks-Carter, Muran, & Safran, 2018). Additionally, therapists often are placed in situations that may distract them from focusing specifically on their therapist-client relationship; this may explain their lack of attunement to their clients' perceptions of therapy. Examples of this may include therapists who worry about their highly-stressed clients during a lesser stressed client's session, therapists who are undergoing audits and are preoccupied with their files' completion, or during times when therapists are implementing new techniques and are more focused on doing the technique properly than focusing on the working alliance. Regardless of the reasons contributing to this his lack of awareness (or therapeutic attunement), not addressing strains on the working alliance may lead to damaged client perceptions of therapeutic progress, lower rates of retention, and worse therapy outcomes overall (Friedlander, Escudero, Welmers-van de Poll, & Heatherington, 2018). By developing and testing a measure that assesses fluctuations in the working alliance, it is likely that therapists may be able to become more attuned to strains on the working alliance.

Comparing Daily Reports to Overall Reports

Failing to measure fluctuations in dynamic constructs may impede researchers and practitioners' abilities to formulate effective interventions or treatments. The empirical work found within social psychology (and other literatures) may provide support for this assertion. For example, emotional awareness (EA), defined as a person's ability to be aware of and describe the

emotional state of self and others, is expected to vary over time (Lane & Schwarts, 1987). This conceptualization contrasts with the popular Levels of Emotional Awareness Scale (LEAS; Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990), which conceptualizes EA as a trait and therefore a static (fixed) phenomenon (Versluis et al., 2018). Researchers compared results from the LEAS with ecological momentary assessments of EA. Their findings suggest that comparing daily assessments (fluctuations) of EA levels to broader, overall assessments of EA enhances the field's overall understanding of EA and may lead to better modalities of EA treatment (Versluis et al., 2018). Applying this methodology to the working alliance, the current study seeks to develop and test a measure that assesses working alliance fluctuations by comparing session specific reports from the developed measure to responses from an existing alliance measure, such as the Working Alliance Inventory – Short Revised (WAI-SR; Munder et al., 2010) that elicits more global responses.

There is empirical evidence that variables similar to the working alliance are both conceptualized as static and dynamic. Using the example of EA and the working alliance, both are paradoxically expected to vary over time and yet the way established working alliance measures are currently worded seem to invite static responses from clients. One common item in working alliance measures includes "As a result of *these sessions* I am clearer as to how I might be able to change" (italics inserted; Horvath & Greenberg, 1989; Munder et al., 2010). That statement seems to reflect a general level of task agreement between client and therapist that clients may not feel comfortable, or feel that it is appropriate, to rate lower for just one session. In this example, clients may be pulled to thinking more globally about the working alliance, which may influence how well their responses accurately reflect their perceptions of a particular

session. To this end, this study seeks to develop and test a measure that alters the wording of the most commonly used alliance items to invite more dynamic responses that indicate fluctuations.

The Working Alliance and Outcomes in Therapy

Stronger working alliances are associated with more positive outcomes in therapy (Duncan et al., 2003). Whether these outcomes be a reduction in client symptomology or achieving a goal, researchers have found significant positive correlations between measures of the working alliance and reported outcomes such as on the Outcome Questionnaire (OQ; Taber, Leibert, & Agaskar, 2011). Additionally, Duncan et al. (2003) highlight in their study of over 1,000 research findings that strong working alliances are one of the best predictors of outcomes in therapy. They report conservative average effect sizes of the correlation between the working alliance and therapy outcomes to be at $r = 0.26$. A meta-analysis by Wampold (2001) calculated 54% of the variance in therapy outcomes to be attributed to the working alliance. Other estimates in the literature note the working alliance to account for up to 14.7% of the variance in therapeutic outcomes – a significant amount when compared to other contributing factors to therapy outcomes (Crits-Christoph, Gibbons, Hamilton, Ring-Kurtz, & Gallop, 2011). It is concerning, then, that no scale currently exists to measure fluctuations in working alliance strength that may aid therapists in detecting weakened or weakening working alliances.

Failures to detect weakened or weakening alliances may be due to clients' tendencies to consistently rate the working alliance highly – otherwise known as a ceiling effect (Owen, Reese, Quirk, & Rodolfa, 2013). It is uncertain whether clients consistently give higher ratings based on their authentic perceptions, based on how they think their therapist wants them to answer (termed "social desirability"), responding simply on the therapeutic bond and not the total work/process, or based on fears of a therapist's reaction to honest feedback (Bhattacherjee, 2012). It is also

uncertain whether samples exhibit a selection bias, in that the type of clients who stay in therapy may be individuals who already believe that therapy works and that their therapist knows best. Challenging the authenticity of these consistently high responses by assessing fluctuations in the working alliance may help in identifying which of these rationales are most likely to be associated with ceiling effects. Assessing working alliance fluctuations may also help therapists to better detect the onset of major negative shifts in the working alliance – otherwise known as ruptures.

Ruptures are a phenomenon defined as moments in therapy involving significant amounts of conflict, tension, or misunderstandings between psychotherapists and their clients (Safran & Muran, 2000). This conflict implies low agreement on goals or tasks within therapy and tension implies a strained bond between client and therapist. As such, a rupture may be viewed as the antithesis of a strong working alliance. Preventing rupture formation is important in therapy, as ruptures are highly impactful on treatment outcomes and early client dropout, and are argued to be inevitable over the course of therapy during negotiations between the two parties (Miller-Bottome, Talia, Safran, & Muran, 2018). For example, prevalence rates of ruptures can be estimated from Eubanks, Muran, and Safran's (2018) meta-analysis, which found that 25-84% of clients experienced a rupture during therapy. Although ruptures are not guaranteed to occur, unaddressed ruptures may lead to negative outcomes such as premature termination and client dropout in therapy (Hunsley, Aubry, Verstervelt, & Vito, 1999; Karekla, Konstantinou, Ioannou, Kareklas, & Gloster, 2019). It is likely that consistent assessment of working alliance fluctuations may be used to accurately track the strength of the working alliance and indicate that a rupture has occurred. Ultimately, this study seeks to develop and test a measure that more

greatly identifies fluctuations of client perceptions of the working alliance, and this may be related to perceptions of the session overall.

Client Perceptions of Sessions

A scale that measures fluctuations in the working alliance should be have concurrent validity with other fluctuations that occur in therapy (Bhattacherjee, 2012). Specifically, it is likely that clients' perceptions of a therapy session are associated with their ratings of the strength of the working alliance in the given session. Ways to rate one's session include session effectiveness, session depth, and session comfort - options provided by the established Session Evaluation Questionnaire (SEQ; Stiles et al., 1994). When clients complete an agreed upon goal in session, they may perceive that session to be "effective." An example of this may be a therapist coaching a client to finally achieve their goal of speaking about their traumatic childhood without bursting into tears. Session depth may be related to the working alliance if clients have a strong bond with their therapists that allow the therapists to make insightful comments or challenges to personal topics. Sex therapy is an example of this when clients may resist talking about sex with a professional that they feel is shaming them or does not understand their sexual difficulties (Brandenburg & Bitzer, 2009). This example also illustrates how session comfort may be related to the working alliance. Clients that have stronger bonds with their therapists are likely to feel more comfortable in therapy. Taken together, it is likely that clients' perceptions of items such as the SEQ's session effectiveness, session depth, and session comfort domains are theoretically associated with fluctuations in the working alliance. This study therefore will test the developed measure for concurrent validity with reports from the SEQ.

The Current Study

To explore the validity of the measure developed in this study, we will use Bhattacherjee's (2012) definitions for construct validity and criterion validity. Construct validity refers to the degree to which a measure adequately represents a target underlying construct. Criterion validity refers to the extent that a given measure relates to one or more external criterion based on empirical observations (Bhattacherjee, 2012). A measure is said to have criterion validity when it behaves as expected, given the theory of that construct. The criterion of the measurement's behavior may include how well a measure relates to the construct that it is proposed to measure (convergent validity), how well one measure relates to constructs that are presumed to occur simultaneously (concurrent validity), and how well a measure successfully predicts a future outcome that it is theoretically expected to predict (predictive validity; Bhattacherjee, 2012). By examining the developed measure's convergent, concurrent, and predictive validities, the authors hope to find evidence in favor of supporting the developed measure as a valid means to assess fluctuations in the working alliance.

The authors propose testing the following four hypotheses to validate the developed measure, titled the Session Specific Alliance Measure (SSAM). First, it is hypothesized that the items developed in the current study will reflect a three-factor structure (agreement on goals, tasks, and bond), consistent with current measures of the working alliance (hypothesis 1). Next, we will test for construct validity by proposing that higher ratings of the SSAM will be associated with higher ratings on the established WAI-SR (hypothesis 2). Concurrent validity will be assessed by testing the association between the SSAM and the SEQ. The authors propose that higher ratings on the SSAM will be associated with higher ratings on the SEQ (hypothesis 3). Finally, it is hypothesized that higher ratings on the SSAM will be predictive of positive

therapeutic outcomes, as indicated by lower self-reported ratings of symptomology over time on the Outcome Questionnaire (OQ)- displaying good predictive validity (hypothesis 4).

Methods

Participants

Individuals engaged in psychotherapy were recruited to complete measures related to their experiences in therapy and consisted of a total of 40 participants in the current study. These individuals were recruited from a therapy center in western United States. No formal diagnoses are collected at this center. Eligible participants needed to be 18 years of age or older to be included in the current study. This eligibility is due to distinct differences between alliances with children and alliances with adults (Friedlander et al., 2018). Client demographic information already gathered from the clinic was planned to be added to the data collected from this study's surveys upon completion of collecting our own data. However, the clinic suddenly and unexpectedly closed, which we noted below as a limitation for obtaining client demographics along with other data. Given the center's history of clientele, it was predicted for this study that participants would be predominantly White/Caucasian, have ages ranging from 18-40 years old, as well as have at least a high-school level of education.

Procedure

After obtaining IRB approval, clients that gave signed consent were invited to participate at a western United States university counseling center. Surveys containing the WAI-SR, the SSAM, the SEQ, and the OQ were administered via university-provided iPads, were handed out by clinic personnel, and was collected upon completion of the surveys for each client. Given the focus on current therapy experiences, clients who have recently terminated therapy were not asked to participate in the current study. Consenting clients were invited to complete the surveys

two weeks following their first survey. This procedure would have continued until obtaining at least three waves of data per participant. However, the clinic unexpectedly closed due to the pandemic (COVID-19, 2020). Therefore, only one wave of data from clients was obtained from each client. This means that each client only rated each of the 4 provided measures one time. Recording multiple waves of responses would have provided more longitudinal data and might have better highlighted the SSAM's ability to sensitively track changes in the working alliance. Clients were asked to report how many therapy sessions they have completed at the time the survey is given, as well as how much time has passed since their last session. Therapists were asked to track any clients who may drop out of therapy and report the reasons why those clients left therapy.

Measures

The Working Alliance. The Working Alliance Inventory-Short Revised was used to measure the working alliance (WAI-SR; Munder et al., 2010). Using self-report, the WAI-SR is a 12-item measure of the working alliance and is rated on a five-point scale ranging from 5 (Always) to 1 (Seldom) (Munder et al., 2010). Cronbach's alpha for the WAI-SR was .91, and four items respectively are used to assess the goal, tasks, and bond dimensions. Example items include "___and I collaborate on setting goals for my therapy" (goals), "I feel that the things I do in therapy will help me to accomplish the changes that I want" (tasks), and "___and I respect each other" (bond).

Sensitively Tracking the Working Alliance. The current study used the Session Specific Alliance Measure (SSAM) scale that was developed to assess fluctuations in the working alliance. However, only one wave of data was gathered, so no comparisons between sessions could be made to determine fluctuations. The SSAM contains the WAI-SR's twelve 5-point

items, which allows participants to respond to questions regarding their goals, tasks, and bond with answers ranging from 5 (Always) to 1 (Seldom). However, the SSAM proceeded each item with the stem "*When thinking about **today's** session"* and SSAM items were given in tandem with WAI-SR items to allow participants the opportunity to reflect how the working alliance in that particular session may have differed from the working alliance in general. This strategy was designed to allow clients to be able to report positively about the alliance in general, which may allow them to feel more comfortable reporting less-positively about a particular session. Items reflected language used by the WAI-SR such as "*When thinking about **today's** session,* ___and I collaborate on setting goals for my therapy" (Munder et al., 2010). Cronbach's alpha for the SSAM was .90, as calculated from the only wave of data collected in the current study.

Client Perceptions of Therapy Session. To assess clients' perceptions of their sessions in therapy, the Session Evaluation Questionnaire (SEQ; Stiles et al., 1994) was used. The SEQ's contains multiple indices, including the most common Session Depth and Session Smoothness subscales (Stiles, 2002). The twenty-four 7-point bipolar semantic differential-styled items begin with "This session was" or "Right now I feel." Example dimensions include (respectively) "shallow/deep" and "rough/smooth" or "angry/pleased" and "uncertain/definite." SEQ subscale indices have been reported with good internal consistencies. Cronbach's alpha for the SEQ was .71.

Client Symptomology. The Outcome Questionnaire (OQ) was used in the current study to assess therapy outcomes through the lens of changes in client symptomology over time (Mueller, Lambert, & Burlingame, 1998). Given the single wave of data collected in the current study, however, we were only able to obtain a snapshot of clients' symptomology, versus actual therapy outcomes. The OQ's Cronbach's alpha was .91 and contained 45 items that are grouped into the

subscales of Subjective Distress (25 items), Interpersonal Relationships (11 items), and Social Role (9 items; Mueller et al., 1998). Items are rated on a 5-point scale ranging from 4 (Never) to 0 (Always), with higher scores indicating worse symptomology (e.g., "I get along well with others" or "I find my work/school satisfying").

Results

Reliability and Exploratory Factor Analysis

Results for the current were taken from the single wave of data collected from participants, meaning that each participant completed the SSAM, the WAI-SR, the SEQ, and the OQ one time. The first hypothesis proposed was that the items developed in the current study will reflect a 3-factor structure. To test this hypothesis, an exploratory factor analysis was conducted along with a test for internal consistency reliability. Results revealed a Cronbach's alpha of .90, indicative of good internal consistency, as well as support for a 3-factor solution. This factor loading matches the factor structure of the WAI-SR, supporting hypothesis 1. Eigenvalues were 2.1 (factor 1), 1.4 (factor 2), and 1.2 (factor 3; see Table 2). The maximum likelihood method (canonical) was used to extract the factors and this was followed by a promax (oblique) rotation. Inspection of the scree plot, the residual correlation matrix, the proportion of variance accounted for, chi-square test of residuals, and interpretability of the factors were examined to determine the number of factors to retain for each sub-sample (Loehlin, 1998; Tabachnik & Fidell, 2007). In interpreting the factor pattern, an item was considered to load sufficiently on a given factor if the factor loading was .35 or greater on that factor and less than .35 on the other factors. Items used for each of the goals, tasks, and bond factors respectively included "___ and I collaborate on setting goals for my therapy," "I believe the way we are

working with my problem is correct," and "___ and I respect each other" (see Table 2). This criterion resulted in all items being retained.

Construct, Concurrent, and Predictive Validity

Due to an unexpected national health crisis (COVID-19, 2020), the authors were unable to collect enough data as intended to account for the nested nature of multiple clients seeing the same therapist. There was a total of 14 therapists whose clients (n = 47) participated in the current study. However, the maximum number of clients per therapist were 6, with a range of 1-6 clients per therapist and the average number of clients per therapist equaling 3 to 4 clients (3.36). Multilevel modeling would have been the preferred analysis to account for nesting effects. Not being able to conduct this type of analysis is noted as a limitation, as having a limited sample size of under 50 participants is likely to lead to biased results in multilevel modeling (Maas & Hox, 2005). Therefore, a multiple regression was performed to test the other 3 hypotheses of the current study. The number of sessions was entered as a control variable. Next, the currently established alliance measure (WAI-SR), session rating (SEQ), and client symptomology (OQ-45) were entered as independent variables, and the newly developed measure of alliance (SSAM) was entered as the dependent variable (see tables 3 and 4). Bivariate correlations were also conducted between each of the variables (see Table 5).

Hypotheses 2 proposed that higher ratings of the SSAM will be associated with higher ratings on the WAI-SR. Findings indicated that higher ratings on the SSAM were indeed significantly positively associated with higher ratings WAI-SR, as evidenced by results from the bivariate correlation ($r = .93$; $p < .001$; see Table 5). These findings support the SSAM to have good construct validity, meaning that it seems that the SSAM adequately represents the working alliance construct.

To test hypotheses 3 and 4, that higher ratings on the SSAM would respectively be associated with higher ratings on the SEQ (concurrent validity) and associated with lower ratings of symptomology on the OQ (predictive validity), two separate multiple regressions were conducted. The WAI-SR was included in these multiple regressions to further establish the SSAM's construct validity, in that the SSAM should reflect the kinds of relationships that the WAI-SR is expected to have with the SEQ and the OQ. The WAI-SR and the SSAM first were analyzed for their associations with session ratings, and then tested in their associations with client symptomology (see Tables 3 and 4). Neither the WAI-SR ($\beta = .06$, $t = .61$, $p = .55$) nor the SSAM ($\beta = .03$, $t = .13$, $p = .90$) were able to yield statistically significant results with regard to ratings of the quality of the session (SEQ). Additionally, results failed to support the WAI-SR ($\beta = .54$, $t = 1.27$, $p = .21$) or the SSAM ($\beta = .40$, $t = .95$, $p = .35$) to yield any significant associations with client symptomology (OQ-45). These results suggest that clients' perceptions were not related to how they would rate the working alliance and suggest that these clients' symptomology was not related to the working alliance as well.

Discussion

Despite the existing body of literature on working alliance measures, no measure to date adequately assesses possible fluctuations in the working alliance. This is concerning, as current measures may be eliciting more global responses from clients, which may be masking changes occurring in the working alliance. If these changes in working alliance strength lead to stress on the working alliance, then poor attunement to these strains on the working alliance may lead to ruptures or negative outcomes such as premature termination or client dropout (Hunsley et al., 1999; Karekla et al., 2019). The current study developed and tested the initial reliability and

validity of the Session Specific Alliance Measure (SSAM), which was intended to assess fluctuations in the working alliance over time.

The results demonstrate that the SSAM's data has a strong internal consistency, given its Cronbach's alpha of .90. Meaning, the items in the SSAM seem to reliably reflect clients' perceptions of one construct – the working alliance. Results also support our hypothesis that the SSAM would exhibit a three-factor structure (see Table 2). This factor loading is consistent with existing working alliance measure factor loadings, capturing each of Bordin's (1979) defined subcomponents of goals, tasks, and bond. This is important, as a composite score from the SSAM is therefore likely to validly reflect clients' perceptions of the working alliance.

To further support the validity of the SSAM, it was stated in our second hypothesis that higher scores from the SSAM would be correlated with higher scores from the Working Alliance Inventory – Short Revised (WAI-SR). As expected, the SSAM had a high correlation with the WAI-SR ($r = .93$; $p < .01$; see Table 5). This was by design, given the fact that the SSAM maintained key language from the items on the WAI-SR (see Figure 2). Given this high correlation, the SSAM may be expected to share similar relationships found with the WAI-SR, such as results showing the WAI-SR to have a significant relationship ($r = -.26$; $p < .05$) with the Outcome Questionnaire (OQ; Mueller et al., 1998; McClintock, Anderson, & Petrarca's, 2015). Given these results, it is likely that the SSAM has good construct validity.

Our next hypothesis was that higher scores from the SSAM would be correlated with higher scores from the Session Evaluation Questionnaire (SEQ). It is surprising to note that despite the theoretical association between the working alliance and client perceptions of the session, no significant relationship was found between the SSAM and the SEQ. This finding suggests that how a client perceives a session in therapy may not necessarily be attributable to or

considerate of the working alliance. Meaning, how a client views their working alliance is different than how they view their sessions going. Therapy attendance is not limited to a goal-oriented mentality for clients, which may explain the discrepancy found between the current study's results between the SSAM and the SEQ.

To further explore this unexpected finding, it is possible that clients rate their sessions based more on the emotional experiences resulting from established goals and tasks in therapy, rather than based on the goals and tasks themselves. For example, a client who is working through grieving of a lost loved one is not necessarily going to rate a session based on how well they think they agree with their therapist on the tasks of grieving. The process of grieving, when done in a healthy manner, may be described as overwhelmingly cleansing and refreshing (Prigerson, & Maciejewski, 2008). The client may therefore just want to reflect on those kinds feelings of relief and renewal and enjoy the emotional freedom to feel peace, instead of asking their therapist about the next goal they would like to work on in therapy. In this instance, the focus of a client's perception of a session may be more on the actual experience of feeling peace, rather than them thinking about how the progress of therapy is going.

Another possible explanation for the lack of association between the SSAM and the SEQ may be due to the nature of therapy. Therapy is often designed to unbalance (or change) clients' patterns of behaviors or thoughts in such a way that promotes change towards the clients' goals (McCullough, 2003). For example, a client who is working on their public speaking skills in therapy may have various ways to describe the process. If therapy involves exploring the client's fears and challenges them to practice public speaking, the process of unbalancing may then be described negatively, such as "difficult," "uncomfortable," "challenging," or "stressful." However, moments of progress may be described positively, such as "rewarding," "satisfying,"

or "motivating." A strong working alliance may be present in either of these situations. In this example we see that, depending on the session, a session may be rated negatively even though the working alliance may be strong.

Our fourth hypothesis, that higher ratings on the SSAM would be negatively related to client symptomology, was also not supported. Without multiple waves of data, the current study's intended analysis inevitably shifted from examining the association between the SSAM and therapy outcomes, or changes in symptomology over time, to now examining the association between the SSAM's results and a single time point of clients' reported symptomology. Obtaining symptomology reports from only one session is problematic, due to the fact that sessions in therapy are not necessarily only focused on one set of presenting problems in clients; such is the case with mental health disorders that exist in tandem with substance abuse disorders. One example of this includes therapy that addresses the symptomology of both attention-deficit/hyperactivity disorder (ADHD) and alcoholism (Ohlmeier et al., 2008). Although therapy may at times address symptoms from both the client's ADHD as well as their alcoholism, any given session could focus entirely on one set of symptoms or the other. Additionally, sessions may even be devoted to the effects that the client's symptomology has on their family's functioning (a new, third set of symptoms). It is important, then, to obtain information on changes in symptomology over time, as this would help to assess the SSAM's predictive validity and allow for sessions to sometimes deviate from addressing certain types of client symptomology that the SSAM would predict.

Given the results from the current study's data, it appears that clients' reported symptomology was not related to their perceptions of the working alliance for that session. In addition to therapy addressing multiple sets of symptoms over time, it is possible that clients'

responses from the current study indicated more deeply rooted symptomology. An example of deeply rooted symptomology includes symptoms related to trauma such as withdrawing from others or feelings of emotional numbness (Kubany et al., 2004). These symptoms are seen in typical cognitive trauma therapy, which follows an agenda of 11 prescribed sessions in order to accomplish desired changes in clients' symptomology (Kubany et al., 2004). If clients' responses were based on deeply rooted symptomology such as symptoms from trauma, then only capturing one time point of ratings on the OQ may not adequately reflect changes that are occurring over time.

In addition to improving the current study's sample size to better assess the SSAM's predictive validity, the current study would likely have benefitted from a more rigorous and explicit testing of the SSAM's divergent validity or "discriminant validity", which Zhu (2000) argues is the same thing (see limitations). This type of validity is essentially often defined as the opposite of convergent validity; divergent validity is observed when measures have "non-existing or weak correlations" (Paap, Schrier, & Dijkstra, 2019, p. 1295) or are "unrelated" to one another (Kemer et al., 2019, p. 233). It is interesting to note that Bhattacherjee (2012) chose to not use the terminology of the more statistically dichotomous "divergent validity", but instead refers specifically to "discriminant" validity, defining it as the *degree* to which a measure does not measure (or discriminates from) other constructs that it is not supposed to measure. This distinction in definition thus allows for measures of the same construct (in this case, both the WAI-SR and the SSAM assessing the working alliance) to be tested as to whether they validly measure that construct in meaningfully different ways.

Differentiating between common usages of divergent validity and using Bhattacherjee's (2012) definition of discriminant validity is an important distinction to make in assessing the

SSAM's validity. The aim of the current study was not to determine whether or not the SSAM was adequately capturing the working alliance construct per se, which would elicit testing the association between the working alliance and constructs expected to be unrelated. Rather, the purpose of the current study was to test whether the SSAM could validly assess fluctuations of the working alliance, which elicits testing the *degree* to which the SSAM uniquely relates to constructs as compared to other working alliance measures. The SSAM should be assessing the same basic construct (the working alliance) as existing working alliance measures; but if the fluctuations of the working alliance relate to other constructs in a significantly different way than global ratings of the working alliance, this difference should be reflected in the strengths of those relationships.

Based on the findings that the SSAM and the WAI-SR are so highly correlated ($r = .93$; $p < .001$; see Table 5), it is unlikely that the two measures significantly differ- and that the SSAM has low discriminant validity (discussed further in the Limitations section). This suggests that despite the SSAM's purpose in eliciting more session-specific working alliance ratings, these ratings do not seem to be statistically significantly different than ratings given on the more globally-oriented WAI-SR. Additionally, the SSAM behaved very similarly to the WAI-SR in their relationships with the SEQ and the OQ. Neither the SSAM nor the WAI-SR could predict more than 3% of the variance in either the SEQ or the OQ (see Tables 3 and 4). When comparing the standardized betas from each alliance measure, there also was no large differences (WAI-SR: $\beta = .10$ and .17; SSAM: $\beta = .05$ and .10; see Tables 3 and 4). Taken together, these results also fail to support the SSAM as being significantly different than the WAI-SR. Meaning, it is likely that the SSAM has low discriminant validity and is likely measuring the same construct that the WAI-SR measures.

Limitations and Future Directions

There are a number of limitations that may be affecting the results of the currently study. One prime limitation was the unusual halt in data collection as a result of clinic policy changes that occurred due to the COVID-19 pandemic (2020). In addition to not being able to access the clinic's database for client demographics as intended, we were unable to collect more than one partial wave of data, which had multiple effects on the completion of this study. For example, by not being able to ask all the clinic's clients to participate in therapy and obtaining a sample size of under 50 participants, the anticipated nesting effect (from having multiple clients per therapist) was determined to be not as prevalent (Maas & Hox, 2005). The authors therefore decided to change the proposed multilevel model of analysis to performing a multiple regression model.

Additionally, by not having multiple waves of data, it is likely that there was inadequate data to be able to predict therapeutic outcomes on the OQ. Supporting this is the fact that not even responses from the WAI-SR were significantly related to reports on the OQ. This stands in contrast to previous studies which have found significant relationships between results from these exact measures (McClintock et al., 2015). Replication of this study is warranted in order to verify whether the SSAM can validly predict outcomes in therapy.

An interesting limitation may be due to the unique demographics of the clinic's therapists. All therapists were relatively young, graduate students who were working towards completion of their master's degrees in marriage and family therapy. Younger, less experienced therapists may elicit a more patient and less critical relationship from clients who may otherwise be more critical and demanding with a perceived wise and experienced older therapist. It is possible that these kinds of elicited relationships may impact how clients judge their progress in

therapy (goals and tasks) as well as how willing they are to trust their therapists (the bond). In replication of the current study, obtaining a more diverse sample of therapists may be important in order to obtain generalizable results.

It is important to recognize the high correlation between the SSAM and the WAI-SR (see Table 5). Although it was expected that these two measures be correlated to one another, having a correlation of $r = .93$ may be indicative of multicollinearity, or too much similarity between multiple items in these measures. With such a high correlation, it may be possible eliciting more session-specific working alliance ratings are not significantly different than having clients rate the working alliance overall. This lack of difference may be indicative of low support for the SSAM's discriminant validity. Future studies may benefit from testing this possibility through replication of the current study with more than one partial wave of data.

Unlike the current study, which found no significant relationship between either the working alliance measure and the OQ (likely due to the lack of multiple waves of data; compare with results from McClintock et al., 2015), replicating this study with more longitudinal data may additionally benefit how well the SSAM's discriminant validity is assessed. If the strength of the expected relationship between the SSAM and the OQ *identically* matches the strength of the WAI-SR's expected relationship with the OQ, then a more confident conclusion may be made that the SSAM fails to validly assess the working alliance differently than existing working alliance measures. However, if there is an observed difference in the strength of these two correlations (supporting the SSAM's discriminant validity), this would suggest that one measure may be more effective at predicting outcomes in therapy. This type of testing for discriminant validity may also be performed using other measures that have previously been shown to have significant relationships with the WAI-SR. Some of these measures include the helping alliance

questionnaire (Luborsky et al., 1996), the patient satisfaction questionnaire (Marshall & Hays, 1994), or the client task specific change measure – revised (Watson, Greenberg, Rice, & Gordon, 1996).

Conclusion

To assess the fluctuations of the working alliance, the current study developed and tested the validity of the Session Specific Alliance Measure (SSAM). Although preliminary results indicate good construct validity such as internal consistency, expected factor loadings, and high correlations with other working alliance measures, the current study found low concurrent and predictive validity based on failure to find significant relationships between the SSAM and its theoretically related measures. Future studies should aim to replicate this study by obtaining more longitudinal data as was originally intended, using a more diverse sample of therapists, use the originally proposed multilevel modeling method of analysis, and more explicitly and rigorously assess the SSAM's discriminant validity.

References

Bhattacherjee, A. (2012) *Social science research: Principles, methods, and practices*. Textbook collections. Book 3. Retrieved from http://scholarcommons.usf.edu/oa_textbooks/3

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice, 16*(3), 252–260. https://doi-org.ezproxy2.library.colostate.edu/10.1037/h0085885

Brandenburg, U., & Bitzer, J. (2009). The challenge of talking about sex: The importance of patient–physician interaction. *Maturitas, 63*(2), 124-127.

Crits-Christoph, P., Gibbons, M. B. C., Hamilton, J., Ring-Kurtz, S., & Gallop, R. (2011). The dependability of alliance assessments: The alliance–outcome correlation is larger than you might think. *Journal of consulting and clinical psychology*, 79, 267 – 278.

Duncan, B. L., Miller, S. D., Sparks, J. A., Claud, D. A., Reynolds, L. R., Brown, J., & Johnson, L. D. (2003). The Session Rating Scale: Preliminary psychometric properties of a "working" alliance measure. *Journal of Brief Therapy*, *3*(1), 3-12.

Duncan, B. L. (2012). The partners for change outcome management system (PCOMS): The heart and soul of change project. *Canadian Psychology, 53*, 93–104. doi:10.1037/a0027762

Eubanks, C. F., Muran, J. C., & Safran, J. D. (2018). Alliance rupture repair: A meta-analysis. *Psychotherapy*, *55*(4), 508–519. https://doi-org.ezproxy2.library.colostate.edu/10.1037/pst0000185.supp (Supplemental)

Eubanks-Carter, C., Muran, J. C., & Safran, J. D. (2010). Alliance ruptures and resolution. In J. C. Muran & J. P. Barber (Eds.), *The therapeutic alliance: An evidence-based guide to practice.* (pp. 74–94). The Guilford Press.

Falkenström, F., Ekeblad, A., & Holmqvist, R. (2016). Improvement of the working alliance in one treatment session predicts improvement of depressive symptoms by the next session. *Journal of Consulting and Clinical Psychology, 84*(8), 738–751. https://doi-org.ezproxy2.library.colostate.edu/10.1037/ccp0000119

Friedlander, M. L., Escudero, V., Welmers-van de Poll, M. J., & Heatherington, L. (2018). Meta-analysis of the alliance–outcome relation in couple and family therapy. *Psychotherapy, 55*(4), 356-371. https://doi-org.ezproxy2.library.colostate.edu/10.1037/pst0000161

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

Hunsley, J., Aubry, T. D., Verstervelt, C. M., & Vito, D. (1999). Comparing therapist and client perspectives on reasons for psychotherapy termination. *Psychotherapy: Theory, Research, Practice, Training, 36*(4), 380–388. https://doi.org/10.1037/h0087802

Karekla, M., Konstantinou, P., Ioannou, M., Kareklas, I., & Gloster, A. T. (2019). The Phenomenon of Treatment Dropout, Reasons and Moderators in Acceptance and Commitment Therapy and Other Active Treatments. *Clinical Psychology in Europe, 1*(3), 1-36.

Kemer, G., Moe, J., Johnson, K. F., Goodman-Scott, E., Sunal, Z., & Li, C. (2019). Further validation of the Consultation Skills Scale (CSS): Relationships between consultation skills, supervisory working alliance, and interprofessional collaboration. *Professional Counselor, 9*(3).

Kubany, E. S., Hill, E. E., Owens, J. A., Iannce-Spencer, C., McCaig, M. A., Tremayne, K. J., & Williams, P. L. (2004). Cognitive trauma therapy for battered women with PTSD (CTT-BW). *Journal of Consulting and Clinical Psychology, 72*(1), 3.

Lane, R. D., Quinlan, D. M., Schwartz, G. E., Walker, P. A., & Zeitlin, S. B. (1990). The Levels of Emotional Awareness Scale: A cognitive-developmental measure of emotion. *Journal of personality assessment, 55*(1-2), 124-134.

Lane, R., & Schwartz, G. (1987). Levels of emotional awareness: A cognitive-developmental theory and its application to psychopathology. *American Journal of Psychiatry, 144*, 133-143.

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis*, (ed.) Lawrence Erlbaum Associates. Inc., Mahwah, NJ.

Luborsky, L., Barber, J. P., Siqueland, L., Johnson, S., Najavits, L. M., Frank, A., & Daley, D. (1996). The revised helping alliance questionnaire (HAq-II): psychometric properties. *The Journal of Psychotherapy Practice and Research, 5*(3), 260.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86-92.

Marshall, G. N., & Hays, R. D. (1994). The patient satisfaction questionnaire short-form (PSQ-18).

McClintock, A. S., Anderson, T., & Petrarca, A. (2015). Treatment expectations, alliance, session positivity, and outcome: An investigation of a three-path mediation model. *Journal of Clinical Psychology, 71*(1), 41-49.

McCullough, L. (Ed.). (2003). *Treating affect phobia: A manual for short-term dynamic psychotherapy*. Guilford Press.

McLeod, B. D., & Weisz, J. R. (2005). The therapy process observational coding system-alliance scale: Measure characteristics and prediction of outcome in usual clinical practice. *Journal of Consulting and Clinical Psychology*, *73*(2), 323.

Miller-Bottome, M., Talia, A., Safran, J. D., & Muran, J. C. (2018). Resolving alliance ruptures from an attachment-informed perspective. *Psychoanalytic Psychology, 35*(2), 175.

Mueller, R. M., Lambert, M. J., & Burlingame, G. M. (1998). Construct validity of the Outcome Questionnaire: A confirmatory factor analysis. *Journal of Personality Assessment, 70*(2), 248-262.

Munder, T., Wilmers, F., Leonhart, R., Linster, H. W., & Barth, J. (2010). Working Alliance Inventory-Short Revised (WAI-SR): Psychometric properties in outpatients and inpatients. *Clinical psychology & psychotherapy, 17*(3), 231-239.

Ohlmeier, M. D., Peters, K., Wildt, B. T. T., Zedler, M., Ziegenbein, M., Wiese, B., ... & Schneider, U. (2008). Comorbidity of alcohol and substance dependence with attention-deficit/hyperactivity disorder (ADHD). *Alcohol & Alcoholism, 43*(3), 300-304.

Owen, J., Reese, R. J., Quirk, K., & Rodolfa, E. (2013). Alliance in Action: A new measure of clients' perceptions of therapists' alliance activity. *Psychotherapy Research, 23*(1), 67–77. https://doi-org.ezproxy2.library.colostate.edu/10.1080/10503307.2012.731088

Paap, D., Schrier, E., & Dijkstra, P. U. (2019). Development and validation of the Working Alliance Inventory Dutch version for use in rehabilitation setting. *Physiotherapy Theory and Practice, 35*(12), 1292-1303.

Prigerson, H. G., & Maciejewski, P. K. (2008). Grief and acceptance as opposite sides of the same coin: setting a research agenda to study peaceful acceptance of loss. *The British Journal of Psychiatry, 193*(6), 435-437.

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

Stiles, W. B., Reynolds, S., Hardy, G. E., Rees, A., Barkham, M., & Shapiro, D. A. (1994). Evaluation and description of psychotherapy sessions by clients using the Session Evaluation Questionnaire and the Session Impacts Scale. *Journal of Counseling Psychology*, *41*(2), 175–185. https://doi-org.ezproxy2.library.colostate.edu/10.1037/0022-0167.41.2.175

Stiles, W. B. (2002). Session evaluation questionnaire: Structure and use. *Journal of Clinical Psychology*, *55*, 10-12.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics, 4th ed.* Allyn & Bacon/Pearson Education.

Taber, B. J., Leibert, T. W., & Agaskar, V. R. (2011). Relationships among client–therapist personality congruence, working alliance, and therapeutic outcome. *Psychotherapy, 48*(4), 376.

Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings.* Mahwah, NJ: Erlbaum.

Watson, J. C., Greenberg, L. S., Rice, L. N., & Gordon, L. B. (1996). *Client Task-Specific Change Measure-Revised.* Unpublished measure, University of Toronto, Canada.

Zhu, W. (2000). Which should it be called: Convergent validity or discriminant validity? *Research Quarterly for Exercise and Sport, 71*(2), 190-194. Doi:10.1080/02701367.2000.10608897

Appendix

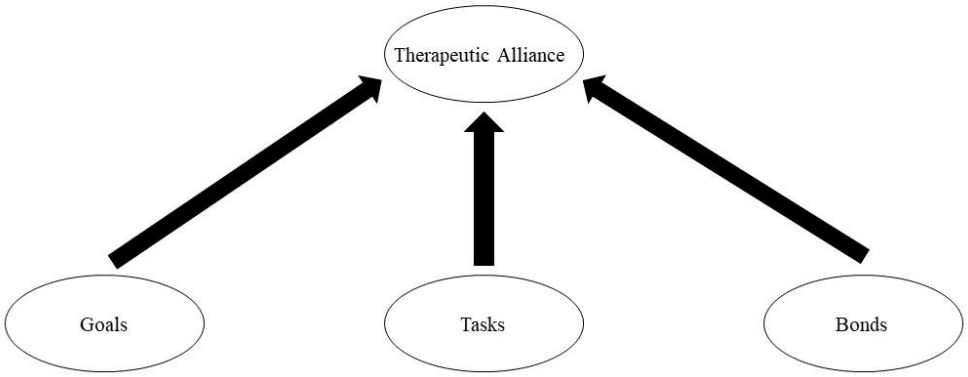

Figure 1. *Conceptual Model of the Components of the Therapeutic Alliance*

(WAI-SR)

_____and I are working towards mutually agreed upon goals.

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

…

(SSAM)

*When thinking about **today's** session…* _____and I are working towards mutually agreed upon goals.

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

Figure 2. *Example Items from the WAI-SR and SSAM*

Table 1

Descriptive Statistics

Scale	Range	Min - Max	Mean	Standard Deviation
WAI-SR	1-5	2.25 – 5.00	4.45	.58
SSAM	1-5	2.25 – 5	4.55	.55
SEQ	1-7	2.90 – 4.05	3.58	.28
OQ	1-5	1.36 – 3.31	2.03	.41

Table 2

SSAM Exploratory Factor Analysis

Item	F1	F2	F3
1. As a result of these sessions, I am clearer as to how I might be able to change.	**.69**	.33	.41
2. What I am doing in therapy gives me new ways of looking at my problem.	**.73**	.37	.32
3. I believe the way we are working with my problem is correct.	**.82**	.27	.38
4. I feel that the things I do in therapy will help me to accomplish the changes that I want.	**.72**	.30	.40
5. ___ and I collaborate on setting goals for my therapy.	.44	**.76**	.40
6. ___ and I are working towards mutually agreed upon goals.	.14	**.84**	.41
7. ___ and I have established a good understanding of the kind of changes that would be good for me.	.71	**.73**	.22
8. ___ and I agree on what is important for me to work on.	.43	**.82**	.23
9. I feel that ___ appreciates me.	.23	.40	**.79**
10. I feel ___ cares about me even when I do things that he/she does not approve of.	.45	.38	**.81**
11. ___ and I respect each other.	.31	.33	**.78**
12. I believe ___ likes me.	.55	.43	**.87**

Table 3

Regressions with SEQ as Dependent Variable

	B	*SE B*	*β*	*t*	*p*	*R Squared*
WAI-SR	.06	.09	.10	.61	.55	.01
SSAM	.03	.20	.05	.13	.90	.03

Table 4

Regressions with OQ as Dependent Variable

	B	*SE B*	*β*	*t*	*p*	*R Squared*
WAI-SR	.17	.12	.17	1.04	.31	.03
SSAM	.07	.12	.10	.60	.57	.01

Table 5

Bivariate Correlations Between Measures

Variable	1	2	3	4
1. WAI-SR	-			
2. SSAM	.93**	-		
3. SEQ	.19	.20	-	
4. OQ	.17	.10	.21	-

**. Correlation is significant at the 0.01 level (2 – tailed)

*. Correlation is significant at the 0.05 level (2 – tailed)