

**THE LONGITUDINAL TRAJECTORY AND CLIENT-THERAPIST
AGREEMENT OF PERSONALITY TRAITS OVER THE COURSE OF
THERAPY**

by

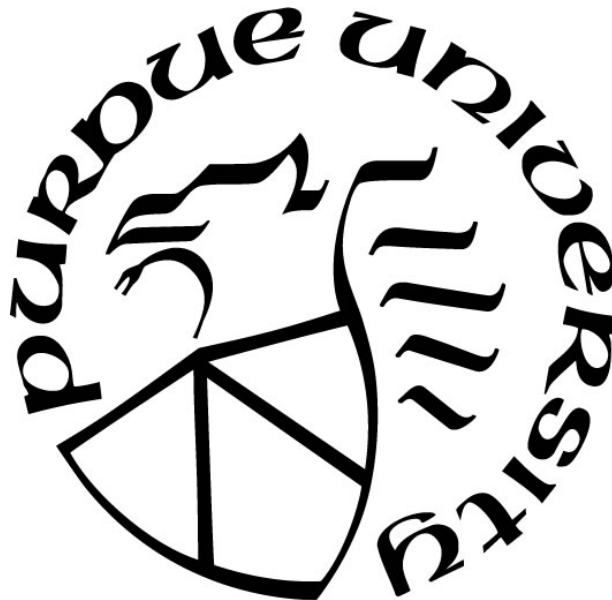
Meredith A. Bucher

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Department of Psychological Sciences

West Lafayette, Indiana

August 2022

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Douglas B. Samuel, Chair

Department of Psychological Sciences

Dr. Donald R. Lynam

Department of Psychological Sciences

Dr. Brent W. Roberts

Department of Psychology, University of Illinois

Dr. Susan C. South

Department of Psychological Sciences

Approved by:

Dr. Kimberly P. Kinzig

To my advisor, Douglas Samuel, whose scientific curiosity sparked my own passion for lifelong learning greater than I could have ever imagined,

To my husband, Korey Bucher, for being my biggest cheerleader and support system,

To my daughters, Brynn and Autumn Bucher, for their patience and love as I navigated motherhood as a graduate student,

To my parents, Myra and Douglas Cors, who taught me at a young age that no goal is unreachable,

And to my cat, Bodhi, who provided companionship during every late night spent on my graduate work.

ACKNOWLEDGMENTS

First and foremost, I thank my advisor and dissertation chair, Dr. Douglas Samuel, who took a chance on me six years ago and has provided invaluable guidance in research, teaching, and mentorship ever since. Additionally, I thank my committee members, Drs. Donald Lynam, Brent Roberts, and Susan South, for the insightful feedback and support that made this study stronger. I also thank Dr. Sean Lane who provided feedback and assistance regarding statistical analyses, primarily those related to the MLM power analyses. I thank Purdue University's clinical psychology department for the plethora of support provided during my tenure here. Last but certainly not least, I thank the many community members who entrusted the Purdue Psychological Treatment and Research Clinics to enhance their lives and provide hope when they needed it most.

TABLE OF CONTENTS

LIST OF TABLES	7
LIST OF FIGURES	8
ABSTRACT	9
INTRODUCTION	10
Client and Therapist Report of Personality.....	11
Client and Therapist Agreement	13
Research Questions and Hypotheses	15
METHOD	17
Client and Therapist Information.....	17
Measures	17
Five-Factor Model Rating Form.....	17
Personality Assessment Inventory	20
Clinical Outcomes in Routine Evaluation	20
Treatment and Rating Details	21
Analyses	22
Assessing Personality Change Across Time and Raters.....	22
Comparison of Client-Therapist Agreement Across Sessions.....	23
Client-Therapist Agreement and its Relationship With Therapeutic Outcomes	27
RESULTS	28
Client-Therapist Change Across Sessions	28
Client-Therapist Agreement Across Sessions.....	42
Agreement as Predictors of Engagement and Improvement.....	51
Supplementary, Exploratory, and Post-hoc Analyses.....	55
DISCUSSION	58
Personality Change Across Time.....	58
Client-Therapist Agreement Across Time	60
Client-Therapist Agreement as Predictors of Treatment Outcomes	64
Clinical Implications	65
Limitations and Future Directions	66

Summary	67
REFERENCES	69

LIST OF TABLES

Table 1. Client Demographics	18
Table 2. Assessment Rating Information.....	25
Table 3. Domain Means and Standard Deviations for Each Session for All Clients and Therapists	29
Table 4. Domain Means and Standard Deviations for Each Rating for a Subset Sample	34
Table 5. Linear Mixed-Effects Model of Client and Therapist Personality Ratings Over Time, not Including Distress	37
Table 6. Linear Mixed-Effects Model of Client and Therapist Personality Ratings Over Time, Including Distress as a Covariate.....	38
Table 7. Correlations Between Therapist and Client Ratings.....	44
Table 8. Partial Correlations Between Therapist and Client Ratings, Controlling for Distress ...	45
Table 9. Paired Samples T-tests of Client and Therapist Ratings.....	46
Table 10. Average ICCs for Each Rating	48
Table 11. Average ICCs for Thirty Client-Therapist Dyads During the First Four Ratings	48
Table 12. Linear Mixed-Effects Model of Client-Therapist Agreement	49
Table 13. Correlations Among Improvement, Engagement, and Client-Therapist Agreement ...	52
Table 14. Linear Regression of Agreement Predicting Improvement	53
Table 15. Linear Regressions of Agreement Predicting Attendance and Engagement	54
Table 16. Facet-Level Scores per Rating for Clients and Therapists.	56

LIST OF FIGURES

Figure 1. Mean-level personality trait trends for client and therapist ratings at each assessment.	32
Figure 2. Spaghetti plot of all client and therapist ratings across days used in analyses.....	33
Figure 3. Mean-level personality trait trends for a subset ($n = 50$) of clients through the first four ratings.....	35
Figure 4. Spaghetti plots for a subset ($n = 50$) clients that at least completed up to the fourth rating.	36
Figure 5. Therapist- and client-reported neuroticism plotted with client's ratings of distress throughout treatment.	40
Figure 6. An overlay of the first four personality ratings for the entire sample, a subset of 50 with the first four ratings, and those with ratings at each time point that included both dyads.....	41
Figure 7. Plotted agreement (ICCs) across the first four assessment time points.....	50

ABSTRACT

Personality traits are important factors of psychotherapy for many reasons, as they relate to a variety of clinical outcomes, can complicate treatment, and can also be targets of treatment interventions. Because of its clinical prevalence and impact, it is imperative that therapists are able to effectively assess and treat personality pathology. Previous research has indicated that both client and therapist ratings of personality can provide meaningful information, and this varies across different sessions, but no study to date has examined *both* client and therapist ratings across the entire therapeutic intervention. There is also limited information on the agreement of client and therapist ratings of personality, as the majority of studies only examine the outset, the end, or a random time point of treatment. Examining only one point in time – or just the beginning and end – misses valuable information regarding possible changes in personality occurring throughout treatment. Using a naturalistic dataset of 128 client-therapist dyads (3,440 observations), the present study examined the longitudinal trajectory of client and therapist ratings of personality change throughout intervention while also accounting for state-level distress. Additionally, the agreement between clients and therapists were examined throughout treatment for any potential patterns of change using rank-order, mean-level, and absolute agreement. Significant patterns of trait change and change in absolute agreement across treatment were assessed using multilevel modeling. Last but not least, the agreement among clients and therapists were examined as potential predictors of therapeutic outcomes, such as engagement and improvement. The results provided evidence for significant decreases in neuroticism that were reported by the client but not therapist that suggest clients might be report decreases in state-level distress rather than true trait change. There were meaningful fluctuations in agreement across treatment, particularly for openness to experience and neuroticism, but the overall agreement – or lack thereof – did not significantly predict client engagement or improvement. Results highlight several clinical implications that are discussed.

INTRODUCTION

Personality pathology has great clinical and public health significance. A review from Samuels (2011) suggests 5-10% of the general population could be diagnosed with a personality disorder, although this is likely under-representative of its actual prevalence. Personality pathology has been shown to strongly correlate with many common mental health disorders, such as depression and anxiety, despite being relatively under-assessed in clinical mental health settings (Newton-Howes et al., 2010). Personality traits relate to a variety of clinical outcomes, such as therapeutic alliance, abstinence, and overall improvement (Bucher, Suzuki, & Samuel, 2019), and it often complicates the treatment of mental health disorders. Those with greater personality pathology tend to utilize and need acute therapeutic services, such as inpatient treatments, more frequently (Bateman & Fonagy, 2008) and often have higher rates of drop out (Crawford et al., 2007; Swift & Greenberg, 2012). Personality pathology can also impact therapist performance, as it is associated with clinician burnout as well (Linehan, Cochran, Mar, Levensky, & Comtois, 2000). Taken together, its clinical prevalence and impact calls for a great need to effectively assess and treat personality pathology in clinical settings.

Although personality was once considered relatively fixed and unmalleable, research has shown otherwise (e.g., Roberts & Mroczek, 2008; Soto, John, Gosling, & Potter, 2011). Personality trait change is often conceptualized from a behavioral perspective (e.g., Allemand & Flückiger, 2017; Wrzus & Roberts, 2016), such that specific events might result in new behavioral responses, or immediate states. Over time, increases in the frequency of any new specific state can result in long-term behavioral, or personality trait, change. Personality trait change is associated with a variety of life outcomes related to physical health and well-being (Allen, Vella, & Laborde, 2015; Human et al., 2013; Turiano et al., 2012), highlighting the importance of examining specific patterns of trait change as well as environments or triggers that can lead to change. A systematic review conducted by Roberts and colleagues (2017) found that in one such environment, the therapeutic setting, clients showed significant increases in levels of emotional stability, extraversion, openness, and agreeableness in individuals seeking psychological treatment. Not only that, but personality changed rather quickly, often occurring within the first month of treatment. A caveat to these findings, however, is that these ratings were exclusively self-report. Thus, although the clients reported changes on traits, informants such as close friends and family

or therapists might not. Second, these analyses were unable to differentiate state-level change from trait-level change. That is, the personality trait change might have rather been due to decreases in state-level distress, a concept described as the *state-artifact position*. Nonetheless, these findings challenge the belief that personality traits – including maladaptive levels of personality traits – cannot be targeted as or changed with interventions. In fact, although few have examined this, recent studies have shown interventions targeting personality do appear to result in trait change, ranging from structured to unstructured protocols (Allan, Leeson, Fruyt, & Martin, 2018; Hudson & Fraley, 2015; Sauer-Zavala, Fournier, Steele, & Woods, 2020).

Client and Therapist Report of Personality

In order to target personality as a treatment intervention, clinicians must first assess clients' personality function as they would any other psychological symptom of clinical interest. For many reasons, clinical personality assessments often rely primarily on the perspective of the client. One reason might be related to overall efficiency as clients might not return informant-report measures or bring informants to sessions, and therapists might be hesitant to complete personality assessments themselves due to lack of reimbursement from insurance companies (Eisman & Nordal, 2017). Therapists might also believe that because clients know themselves best, including their own thoughts, beliefs, and behaviors, that clients would provide the most information about their personality resulting in little need for obtaining information from an informant. Yet, research by Samuel, Bucher, and Suzuki (2018) showed that both clients' *and* therapists' reports of the client's personality related to various outcomes in meaningful ways, such as engagement in treatment and symptom improvement. Not only that, but Samuel and colleagues found that these associations changed whether the reports were at the first or fourth session, suggesting both clients and therapists can contribute important clinical information across therapeutic services. This provides support for therapists as valuable sources for obtaining information on clients' personality functioning as they possess information incremental to clients' reports across psychotherapy.

The therapeutic relationship is unique from most other relationships for a variety of reasons, one of which is the clients' disclosure of intimate details, including mental health, after having known the therapist for a limited amount of time. Because of this, therapists receive a great deal of rich information from client interactions, and they are likely exposed to clients' thoughts to which friends or family might not have frequent access. This might be particularly beneficial for

therapist-ratings of traits with low observability such as neuroticism. Additionally, therapists are presumably trained to identify and interpret information related to personality, behavior, and their impact on various areas of client functioning. Thus, therapists might recognize subtle patterns of thoughts, emotions, and behaviors before an acquaintance or friend. Yet, a meta-analytic investigation by Samuel (2015) suggested that the agreement between clients and therapists (*mdn* $\kappa = .11$) was actually somewhat lower than that generally found between self-report and spouses or peers. A caveat to this finding, however, is the methods of obtaining information were not the same across therapists and clients, and research has consistently found the agreement across different measures and methods to be quite poor across all psychopathology ($r = .30$; Achenbach, Krukowski, Dumenci, & Ivanova, 2005).

Trait agreement – primarily agreement between client and therapist – has largely been examined by rank-order agreement, or correlational analyses. In other words, rank-order agreement between two informants reflects whether the ranking of a given individual is relatively similar across two raters. Higher correlation coefficients imply that the higher a target (i.e., client) rates themselves on a given trait, the higher an informant (i.e., therapist) will rate the target on that same trait. While this approach has provided a great deal of information, a limitation is that it does not allow for the direct comparison of mean-level differences. In order to examine mean-level differences of personality, the same measure must be administered to both the target and informant, an approach that was once difficult to do due to the lack of similarity between client- and informant-reported measures and methods. Using the same measure for both clients and therapists (the Personality Inventory for DSM-5, or PID-5; Krueger, Derringer, Markon, Watson, & Skodol, 2012), Samuel, Suzuki, Bucher, and Griffin (2018) found the agreement to be the same as, and in some cases higher than, self-other agreement with friends and family, further suggesting therapists *are* important sources of client information and should be examined.

Samuel and colleagues were also one of the first to examine specific mean-level differences of reported personality pathology between therapists and clients, providing insight into not only how well clients and therapists agreed with one another but who tended to report more personality pathology. Somewhat surprisingly, clients tended to report significantly higher levels of maladaptive personality traits compared to their therapists. While this provided a great deal of information on clients' ratings compared to their therapists', a limitation to this study was that mean-level differences were not examined over time and across sessions. In fact, research

examining longitudinal personality change over time as reported by the self *and* an additional informant is few, and there is no study to date that has examined this between therapists and clients across treatment with more than two assessment points. As a result, little is known about how therapist-client agreement might change across treatment sessions.

Client and Therapist Agreement

A meta-analysis conducted by Connelly and Ones (2010) found differences in levels of agreement between self- and informant-reports based on specific personality traits. Observable traits, such as those related to extraversion, are more easily seen and evaluated whereas other traits, such as those related to neuroticism, consist of more internal processes not as easily available for informants to perceive, consistent with previous research investigating self- and other-reports of personality (Vazire, 2010). Consequently, self-other agreement tends to be higher for more observable traits compared to less observable traits in most relationships. While it might be expected that therapists have more access to information regarding clients' levels of neuroticism, thereby increasing its observability, Samuel, Suzuki, Bucher, and Griffin (2018) actually found that second to traits related to psychoticism, the largest discrepancies between client and therapist-report tended to be traits related to neuroticism, such as depressivity and emotional lability, such that clients tended to report significantly higher amounts of these traits compared to their therapists. The average number of sessions therapists had with clients before making these ratings were 20, but this varied significantly. Examining only one (varied) point in time provides no information on how agreement of a less observable trait such as neuroticism emerges over time, if at all. That is, if a client reports greater amounts of personality pathology compared to their therapist after twenty sessions, it might be that the discrepancy was even greater at the outset of treatment. This might suggest that the discrepancy will continue to lessen as sessions progress. On the other hand, it might be that the discrepancy between the client and therapist remained the same throughout twenty sessions of therapy. If so, this could have negative consequences on various treatment outcomes, because if the client and therapist see the clients' personality functioning differently throughout treatment, therapists might target different treatment goals than clients would like, or they might conceptualize the client incorrectly.

No study to date has investigated mean-level differences and agreement between clients and therapists in a longitudinal sample, or throughout therapeutic services. Agreement across

services could have significant implications, as therapists' knowledge of clients' symptomatology is an important component for utilizing appropriate treatment interventions (Lambert, 2007). In much the same way, with the prevalence of personality pathology in clinical settings, an understanding of clients' actual *and* perceived personality difficulties would aid therapists in utilizing appropriate interventions. For example, Widiger and Presnall (2013) suggested that clients with high levels of neuroticism might benefit from psychopharmacotherapy, and some research utilizing selective serotonin reuptake inhibitors have found promise in this treatment as a target for neuroticism (e.g., Tang et al., 2009). Widiger and Presnall also suggest that those maladaptively high in agreeableness might benefit from assertiveness training, and those high in conscientiousness might benefit from a highly structured treatment, such as cognitive behavioral therapy. Meta-analytic research has also found that agreeableness is positively related to the therapeutic alliance (Bucher et al., 2019), an important clinical outcome given its relationship to dropout in individuals with personality pathology (McMurran, Huband, & Overton, 2010), and conscientiousness positively relates to abstinence. Taking everything into consideration, it is fruitful to conceptualize and target clients' personality pathology in order to provide appropriate interventions and plan for potential barriers to treatment, such as therapy-interfering behaviors related to personality functioning.

Before targeting these traits as clinical interventions, however, it is important to ensure that therapists can accurately assess them. Thus, the proposed study will examine, in a longitudinal sample, mean-level differences and agreement between therapists' and clients' report of clients' personality traits across multiple treatment sessions. In other words, this study will examine whether client personality change across treatment sessions are perceived differently across therapists and clients, and whether these differences might relate to important treatment outcomes such as client retention and improvement. This study will primarily investigate personality via the five-factor model (FFM), which consists of five bi-polar domains labeled extraversion vs. introversion, agreeableness vs. antagonism, conscientiousness vs. disinhibition, neuroticism vs. emotional stability, and openness to experience vs. closedness to experience. Examining a dimensional trait model of personality is increasingly important, as the field has long argued for utilizing dimensions of personality pathology for diagnosis and treatment conceptualization (e.g., Bagby, Gralnick, Al-Dajani, & Uliaszek, 2016; Bernstein, Iscan, & Maser, 2007; Harkness & Lilienfeld, 1997; Widiger & Costa, 1994). The FFM has displayed links to personality pathology

(Samuel & Widiger, 2008), and clinical utility (Glover, Crego, & Widiger, 2012; Mullins-Sweatt & Lengel, 2012; Reynolds & Clark, 2001), highlighting the clinical significance of this model.

Research Questions and Hypotheses

The following study sought to examine a) change in clients' personality traits over the course of therapy from the perspective of both clients and therapists while also taking into account changes in state-level distress, b) change in agreement between clients and therapists over the course of therapy, and c.) whether client-therapist agreement at different points of therapy predicted important outcomes such as client engagement, retention, and symptom improvement. These findings would further elucidate whether trait change within therapy might be more likely explained by changes in state-level distress and add to the literature examining personality assessment, client-therapist agreement, and predictors of treatment outcomes.

The specific research plan, hypotheses, and analyses were registered and can be found via <https://osf.io/4uqh8>. Specific hypotheses were:

- 1a.) Personality trait levels would change across sessions towards more adaptive levels, and this pattern of change would be similar across therapists and clients. Generally, it was expected that neuroticism would decrease while extraversion, agreeableness, conscientiousness, and openness would increase.
- 1b.) These changes would be most evident for neuroticism, such that there would be greater decreases in this domain compared to the increases expected in the other four factors.
- 2.) On average, clients would report greater levels of personality traits compared to therapists at the outset of therapy. This would be most strongly seen for neuroticism, which is indicative of overall distress. Additionally, it was hypothesized that clients who described themselves in a particularly negative manner, would report even greater levels of personality pathology compared to their therapists.
- 3.) Agreement between clients and therapists would increase with subsequent sessions the client attends.
- 4.) While purely exploratory in nature, higher agreement at the outset of therapy would predict greater engagement in therapy, but not necessarily symptom improvement. This was predicted to be true across all five domains.

Changes and mean-level differences across clients' facet-level reports were also examined. However, this was purely exploratory, and I had no hypotheses regarding specific facets.

METHOD

Client and Therapist Information

This study used de-identified data collected from the Purdue Psychology Treatment and Research Clinics (PPTRC), the in-house training clinic for Purdue University's clinical psychology doctoral program that serves the greater Tippecanoe County in Indiana. Clients who received services between October 2011 and March 2020 were included, which is described in more detail below. As of March 2020, 150 clients had received services, 128 of which had at least one confirmed rating of the FFMRF by both therapist and client and were used in the analyses. Table 1 details client demographics. More than half (66%) identified as female, and about three-quarters identified as Caucasian (76%). The mean age was 29. The most common diagnoses, given by therapists after the therapeutic assessment, were depressive (28%) and anxiety (22%) disorders. Nine percent of the sample was given a personality disorder diagnosis or rule-out. The average length of service was about 23 weeks, or slightly over five months, and the average number of sessions attended was around 15. Twenty-eight doctoral students (68% female) in their 2nd year of training or beyond provided therapeutic services under the supervision of a licensed clinical psychologist. Group and individual supervision were provided to the clinicians on a weekly basis.

Measures

Five-Factor Model Rating Form

The Five-Factor Model Rating Form (FFMRF; Mullins-Sweatt, Jamerson, Samuel, Olson, & Widiger, 2006) is a brief, one-page rating form of the FFM that provides one item per face of each domain, resulting in 30 items. The measure was designed to assess the bipolarity of personality facets, such that items are rated on a scale of 1 (*extremely low*) to 5 (*extremely high*). Included with each item are descriptors for each extreme to aid individuals in making their assessments. For example, the 'extremely high' end of the facet depressiveness is described as *pessimistic, glum*, while the 'extremely low' end of depressiveness is described as *optimistic*. Clients and therapists completed the FFMRF at the initial intake session and generally every fourth session thereafter.

Table 1. Client Demographics

Variable	<i>n</i>	%
Self-Reported Sex		
Female	84	66%
Male	44	34%
Age	29	
	(13-72)	
Ethnicity		
African American	6	5%
Asian/Pacific Islander	6	5%
Caucasian	97	76%
Hispanic/Mexican	8	6%
Native American	1	1%
Multiracial	6	5%
No information reported	3	2%
Marital Status		
Single, never married	80	63%
Single, divorced	13	10%
Engaged or cohabitating	6	5%
Married	25	20%
No information provided	4	3%
Education		
Currently in high school	26	20%
High school diploma	10	8%
Some college/currently in college	39	30%
Bachelor's degree	21	16%
Master's degree/in graduate school	17	13%
Doctorate	3	2%
No information provided	12	9%

Table 1 continued

Variable	<i>n</i>	%
Employment		
Employed	48	38%
Student	43	34%
Student and employed	11	9%
Unemployed	14	11%
Stay-at-home caregiver	2	2%
No information provided	10	8%
Diagnostic Impression		
Attention deficit/hyperactivity disorder	15	11%
Autism spectrum disorder	5	4%
Schizoaffective disorder	1	1%
Bipolar disorder	6	4%
Depressive disorder	38	28%
Anxiety disorder	30	22%
Obsessive-compulsive and related disorder	5	4%
Trauma- and stressor-related disorder	15	11%
Feeding/eating disorder	1	1%
Substance-related disorder	9	7%
Personality disorder	13	9%
Length of service (weeks)	22.89 (1-131)	
Number of sessions	15.11 (1-98)	
Sessions canceled, rescheduled, or no-showed	3.85 (0-17)	21% (0%-78%)

Table 1 continued

Variable	<i>n</i>	%
RCI	-0.68 (-3.29-2.55)	
NIM <i>T</i>-Score	57.09	
Number of personality ratings	<u>Clients</u>	<u>Therapists</u>
	3.44	3.19
	(1-13)	(1-13)

Personality Assessment Inventory

The Personality Assessment Inventory (PAI; Morey, 2007) is a 344-item self-report questionnaire that assesses adult psychopathology. Item responses are on a four-point scale and range from *false* to *very true*. The PAI consists of four validity scales, 11 clinical scales, five treatment scales, and two interpersonal scales that allow clinicians to investigate not only psychological symptoms, but other factors related to interpersonal functioning and treatment planning. It also allows for the investigation of various response styles. For this study, the negative impression management (NIM) scale was used to investigate how negative response styles influenced client-therapist agreement.

Clinical Outcomes in Routine Evaluation

The Clinical Outcomes in Routine Evaluation (CORE; Evans et al., 2000) is a 34-item measure assessing various aspects of functioning in clients. There are four subscales that assess subjective wellbeing, symptoms, functioning, and risk/harm. Three of the subscales (wellbeing, symptoms, and functioning) were used to investigate and control for state-level distress when examining personality change over time.

Treatment and Rating Details

Data used in this study were obtained from naturalistic, clinical case series. Clients sought therapeutic services through the PPTRC for various psychological concerns, and upon entering treatment they signed a consent form that provided them with information regarding the use of their data for research purposes. There were no limits to the number of sessions or weeks clients could receive services.

Starting in the fall of 2011, upon entering treatment, clients attended a 90-minute intake session in which both the client and therapist completed the FFMRF and CORE. Clients typically completed these measures prior to the start of sessions and therapists completed them after. Data collected from between Fall of 2011 and Fall of 2013 were obtained via pencil-and-paper, after which all clients completed these measures electronically via a tablet provided in the clinic. In some instances, clients started treatment by completing pencil-and-paper versions and finished services while completing measures on the tablet. After the intake session, the first two to five sessions were spent completing a therapeutic assessment, including the PAI. Throughout treatment services, both therapists and clients completed the FFMRF approximately every 4th session (i.e., Sessions 1, 5, 9, 13, etc.) and the CORE was completed by the client at every session attended. However, it is important to note that, due to the naturalistic nature of the data, some client and therapist ratings were completed at different sessions than would be expected. The average number of FFMRF ratings completed across the 128 clients was slightly over three for both clients and therapists (see Table 1 for ranges), and the modal number was one. The average number of ratings in which both the client and therapist completed the FFMRF together was also three, and the maximum was 12. It is also worth noting that there were six clients who were transferred to another therapist after having received therapy from their primary clinician. Although information regarding all sessions is provided, analyses described below included only the first four time points due to significant attrition after that rating.

In the middle of March 2020, due to the COVID-19 pandemic, the clinic began operating entirely remote. Because of this, the clinic is no longer collecting data on the FFMRF or CORE while using telehealth services. Thus, this study used data that had been collected between October 2011 to March 2020. Also of note, as of August 2020, five clients included in the current analyses were still receiving services from the clinic.

Analyses

Assessing Personality Change Across Time and Raters

The first goal was to examine client personality change across time and raters. To do this, personality domain scores were calculated by averaging the six facet-level items for each respective domain. The date of the ratings was subtracted from the date of the intake session to calculate the days since intake for each client and therapist rating. It is worth noting that there were two client and two therapist ratings in which dates were not able to be calculated based on information provided in clients' paper files. These were coded as missing data. Distress was calculated by averaging items across the CORE's subscales subjective wellbeing, symptoms, and functioning.

Personality change across time and raters was examined using a linear mixed-effects hierarchical model, in which change in personality trait scores were modeled over time (i.e., days since the intake session) for each domain and rater separately. As stated previously, this was used for only the first four rating time points (i.e., around slightly over five months of treatment). At the first level, the model predicted the personality ratings from the intercept and time. Analyses were conducted separately for each domain and rater, and analyses were also conducted that either included distress (i.e., CORE) as a covariate or did not include distress. This model also examined the random slope model, allowing the slope to differ for each client and therapist rating. Neuroticism was reverse scored as emotional stability to aid with interpretation of coefficients.

$$\text{Level 1: score}_{ij} = \beta_0 + \beta_1(\text{time}) + \beta_2(\text{CORE}) + e$$

$$\text{Level 2 intercept: } \beta_0 + \gamma_{00} + \mu_0$$

$$\text{Level 2 slope: } \beta_1 = \gamma_{10} + \mu_1$$

As stated previously, only half of the analyses included CORE, whereas the other half did not. These analyses were conducted in RStudio with package *nlme* (Pinheiro, Bates, DebRoy, & Sarkar, 2021). For the model including both time and distress, sensitivity analyses were conducted in which simulation data were modeled using the parameters of this model. Using a simulation of 1,000 studies, this model was well powered to detect an effect of CORE as low as .13 for neuroticism (.85) and conscientiousness (.83) and an effect as low as .12 for extraversion (.82),

openness (.80), and agreeableness (.85). When examining the power to detect the main effect of time across all domains, sensitivity analyses indicated that the model was well-powered to detect an effect as small as .002 for neuroticism (1.00), openness (1.00), agreeableness (1.00), and conscientiousness, but it would be underpowered to detect an effect as small as .001. The model would be able to detect an effect as small as .0012 for extraversion (.85). The models including therapist-report were well-powered to detect an effect of distress on therapists' ratings that were as low as .14 for neuroticism (.86) and conscientiousness (.87), .13 for extraversion (.83) and openness (.85), and .12 for agreeableness (.81).

Comparison of Client-Therapist Agreement Across Sessions

First, I correlated clients' and therapists' scores for the first four ratings to investigate rank-order agreement. This was done in two separate analyses, one controlling for clients' self-reported levels of distress, and one not controlling for clients' distress. Post-hoc power analyses indicated that, to detect a medium effect for the bivariate correlations, observed power was above .80 for the first two rating timepoints but only .76 for rating 3 and .62 for rating 4. I also examined mean-level comparisons of client and therapist scores using matched-pairs *t*-tests for each rating using SPSS version 26. Post-hoc power analyses conducted in G*Power suggested that with the range of raters at each time point (80-242) and alpha set to $p = .01$ (although the *p*-values are not the primary focus, I set this to .01 to acknowledge the multiple tests conducted), I had an achieved power ranging at or above .98 for all four ratings. Mean difference confidence intervals and Cohen's *d* effect sizes were also examined.

To test absolute agreement among clients and therapists, intraclass correlations (ICCs) were calculated for each paired rating. I examined domain-level agreement by using the six facet-level items by dividing item variance by variance total. Additionally, using all items from the FFMRF, I examined overall agreement across the entire measure. Change in absolute agreement across the full FFMRF was examined over time using another hierarchical mixed-effects model using ICC as the dependent variable. Time was calculated similar to the first analyses. However, in some cases, clients and therapists did not complete the ratings on the same dates. In these instances, time was averaged between the two raters. For example, if a client completed the first rating on the day of the intake (0) and the therapist completed the first rating two days after the intake session (2), the variable was coded as 1. This occurred for 21% of the data used in this

analysis. Number of days between raters was calculated to see if this impacted the agreement (or lack thereof) between clients and therapists, such that it might be possible that clients' and therapists' ratings were significantly different from each other because the therapist was rating the client days after the client rated themselves. As can be seen in Table 2, the average number of days between the two raters for each session was typically between one and two, although in one instance, the difference between two raters was 41 days. Clients' negative self-impressions were calculated by using the *T*-score of the NIM validity scale of the PAI which was grand mean-centered in the analyses.

ICCs were predicted from the intercept and days since the first session (*time*) with two covariates: negative response style (*NIM*) and days between client-therapist ratings (*days*). This model also investigated whether change of therapist throughout services impacted the agreement between client and therapist over time by adding this into the second-level model. Similar to the first set of analyses, I modeled a random slope for time, allowing all slopes to differ across time for dyads.

$$\text{Level 1: } ICC_{ij} = \beta_0 + \beta_1(\text{time}) + \beta_2(\text{NIM}) + \beta_3(\text{days}) + r$$

$$\text{Level 2 intercept: } \beta_0 = \gamma_{00} + \gamma_{01}\text{therapistchange} + \mu_0$$

$$\text{Level 2 slope: } \beta_1 = \gamma_{10} + \gamma_{11}\text{therapistchange} + \mu_1$$

Similar to the above model, a sensitivity analysis was conducted for this model. Using a simulation of 1,000 studies, analyses indicated that this model was well-powered to detect an effect as small as .0008 for time (.82), .005 for negative response style (.94), .02 for days between raters (1.00), and .20 for therapist change (.83).

Table 2. Assessment Rating Information

	Rating 1		Rating 2		Rating 3		Rating 4		Rating 5	
	Clients	Therapists	Clients	Therapists	Clients	Therapists	Clients	Therapists	Clients	Therapists
<i>n</i>	125	124	83-84	79	70	60	50	43	36	35
Session	1.36	1.49	5	5	9	9	14	13	18	17
Modal session	(1-6)	(1-6)	(2-33)	(2-10)	(4-37)	(4-17)	(7-41)	(7-21)	(14-25)	(13-25)
Days since intake	1	1	4	4	8	8	12	12	16	16
Days since intake	4.47	6.02	47.07	46.34	105.79	104.69	159.60	153.93	225.81	211.54
Modal days since intake	(0-71)	(0-71)	(4-286)	(4-184)	(27-350)	(27-350)	(42-397)	(42-388)	(113-557)	(104-571)
Modal days since intake	0	0	28	28	70	63	107	119	155	155
Days between raters	1.59	1.93			1.75			1.86		2.23
Modal days between raters	(0-28)	(0-16)			(0-26)			(0-14)		(0-41)
Distress	0	0			0			0		0
Distress	1.34	1.34	1.13	1.16	.93	.98	.99	.98	1.11	1.06
	Rating 6		Rating 7		Rating 8		Rating 9		Rating 10	
	Clients	Therapists	Clients	Therapists	Clients	Therapists	Clients	Therapists	Clients	Therapists
<i>n</i>	24	24	14	14	10	10	8	6	8	7
Session	22	22	26	26	30	30	34	35	39	39
Modal session	(19-28)	(17-28)	(23-29)	(23-30)	(28-33)	(28-33)	(32-37)	(33-37)	(36-42)	(36-44)
Modal session	20	20	24	24	28	28	32	34	42	36

Table 2 continued

Days since intake	283.71 (164-636)	270.92 (152-643)	326.57 (205-684)	291.93 (194-581)	331.50 (240-411)	423.38 (294-523)	426.67 (309-551)	505.75 (337-733)	511 (337-741)
Modal days since intake	230	195	219	219	337	393	393	511	511
Days between raters	2.50 (0-27)		58 (0-7)		0	7.17 (0-28)		2.14 (0-8)	
Modal days between raters	0		0		0	0		0	
Distress	1.25	1.28	1.08	1.00	.83	1.03	.79	.79	.67
Rating 11									
	<u>Clients</u>	<u>Therapists</u>	<u>Clients</u>	<u>Therapists</u>	<u>Clients</u>	<u>Therapists</u>	Rating 13		
<i>n</i>	5	4	3	2	2	1			
Session	44 (40-51)	44 (40-51)	51 (44-59)	52 (44-59)	52 (48-55)	48			
Modal session	40	40	44	44					
Days since intake	554.80 (384-826)	554.80 (384-826)	560.67 (525-596)	560.50 (525-596)	618 (598-638)	638			
Modal days since intake	554	554	596	596					
Days between raters	0		0		0				
Modal days between raters	0		0		0				
Distress	1.29	1.14	1.05	.00	1.55	1.39			

Client-Therapist Agreement and its Relationship With Therapeutic Outcomes

Full profile ICCs were calculated for each client-therapist dyad which were then used to predict the outcomes of symptom improvement, session engagement, and session attendance. To quantify improvement, the Reliable Change Index (RCI) was calculated for each client by computing the difference between their first and last rating on the CORE and dividing that difference by the standard error of difference. Scores at or below -1.96 indicated reliable decreases in difficulties related to subjective well-being, reported symptoms, and general functioning not likely due to measurement error, whereas scores at or above 1.96 indicate a reliable increase in reported difficulties. Engagement was examined by calculating the percent of missed sessions (on behalf of the client, not due to the therapist) weighted by the number of sessions the client attended. Thus, a client who attended 100% of 15 sessions scheduled would have a higher score in engagement than those who attended only one session without rescheduling, cancelling, or no-showing. This was different from what I had originally proposed, which was the percentage of sessions missed without taking into account number of sessions. I did, however, include unweighted percentage of sessions attended in a separate analysis as proposed, and I also included an analysis investigating general number of sessions attended, which could also potentially be considered a component of client engagement on its own.

Separate multiple regression analyses were conducted for each agreement coefficient up to the fourth rating. Post-hoc power analyses conducted by G*Power indicated that each regression analysis was powered at .84 to detect a medium effect with the full sample of 128 dyads. However, it is important to note that, this is likely lower in subsequent time points due to attrition in the sample.

RESULTS

Table 2 provides information for each FFMRF rating, including the number of clients and therapists who have completed each rating, the average session that each rating was completed, and the average number of days between paired raters.

Client-Therapist Change Across Sessions

Means and standard deviations for all clients and therapists across all ratings are included in Table 3 (see also Figures 1 and 2). Because of the amount of attrition at each rating, resulting in questions whether generalizability of changes in scores across time were due to actual changes or due to attrition, the patterns of changes were also reported for the first four FFMRF assessment points (i.e., up to around 22 weeks/5 months into treatment) for a specific subset of clients ($n=50$) that had completed up to the fourth rating (Table 4; Figures 3 and 4). Figure 6 shows an overlay of scores across time for the entire sample, dyads only (i.e., those that had a complete client-therapist dyad), and the subset mentioned previously. Similar trends are seen across all three groups, increasing the confidence in speaking to the overall pattern of changes.

Results from the mixed-effects model indicated first and foremost that, when clients' distress levels were not included in the model, (Table 5), clients reported significant increase in emotional stability across time and therapists reported significant decreases in clients' levels of conscientiousness. When including distress within the model (Table 6), distress significantly predicted clients' and therapists' ratings of clients' emotional stability and extraversion. That is, the more distress the client reported, the lower the client and therapist described the client in emotional stability and extraversion. In fact, when plotting mean scores of clients and therapists across all time points (Figure 5), one can see a similar pattern in increases and decreases of neuroticism and client-reported distress, particularly for client-ratings of neuroticism. Within the same mixed-effects model that controlled for clients' distress, clients did not report any significant linear changes in their personality trait scores across any domain. Thus, by including distress into the model, clients no longer reported significant increases in emotional stability. Therapists, however, continued to report significant decreases in clients' levels of conscientiousness across time.

Table 3. Domain Means and Standard Deviations for Each Session for All Clients and Therapists

Domain	<u>Rating 1</u>			<u>Rating 2</u>		
	<i>n</i>	Client	<i>n</i>	Therapist	<i>n</i>	Therapist
Neuroticism	126	3.08 (.77)	123	3.15 (.59)	84	2.93 (.74)
Extraversion	126	3.02 (.72)	123	2.98 (.79)	84	3.04 (.72)
Openness	126	3.44 (.71)	123	2.96 (.61)	84	3.40 (.75)
Agreeableness	126	3.67 (.63)	123	3.61 (.53)	84	3.72 (.61)
Conscientiousness	126	3.49 (.81)	123	3.45 (.78)	83	3.53 (.80)
	<u>Rating 3</u>			<u>Rating 4</u>		
	<i>n</i>	Client	<i>n</i>	Therapist	<i>n</i>	Therapist
Neuroticism	70	2.76 (.74)	60	3.15 (.73)	50	2.87 (.73)
Extraversion	70	3.09 (.66)	60	3.00 (.68)	50	3.08 (.79)
Openness	70	3.46 (.67)	60	2.91 (.74)	50	3.46 (.67)
Agreeableness	70	3.75 (.66)	60	3.49 (.57)	50	3.64 (.63)
Conscientiousness	70	3.56 (.85)	60	3.34 (.81)	50	3.45 (.81)
	<u>Rating 5</u>			<u>Rating 6</u>		
	<i>n</i>	Client	<i>n</i>	Therapist	<i>n</i>	Therapist
Neuroticism	36	3.02 (.69)	35	3.15 (.67)	24	3.03 (.85)
Extraversion	36	3.06 (.77)	35	3.00 (.64)	24	3.06 (.65)
Openness	36	3.53 (.69)	35	2.91 (.85)	24	3.47 (.58)

Table 3 continued

Agreeableness	36	3.68 (.79)	35	3.49 (.48)	24	3.61 (.58)	24	3.64 (.56)
Conscientiousness	36	3.50 (.73)	35	3.34 (.61)	24	3.44 (.71)	24	3.18 (.64)
<u>Rating 7</u>								
Domain	<i>n</i>	Client	<i>n</i>	Therapist	<i>n</i>	Client	<i>n</i>	Therapist
Neuroticism	14	2.98 (.73)	14	2.90 (.56)	10	2.80 (.80)	10	3.12 (.68)
Extraversion	14	3.39 (.80)	14	3.07 (.74)	10	3.42 (.85)	10	3.10 (.55)
Openness	14	3.79 (.54)	14	3.06 (.76)	10	3.80 (.48)	10	3.13 (.73)
Agreeableness	14	4.01 (.80)	14	3.75 (.55)	10	3.78 (.50)	10	3.70 (.55)
Conscientiousness	14	3.74 (.83)	14	3.23 (.77)	10	3.88 (.54)	10	3.40 (.68)
<u>Rating 9</u>								
Domain	<i>n</i>	Client	<i>n</i>	Therapist	<i>n</i>	Client	<i>n</i>	Therapist
Neuroticism	8	2.71 (.54)	6	3.31 (.58)	8	2.73 (.65)	7	3.17 (.70)
Extraversion	8	3.04 (.82)	6	2.83 (.54)	8	3.23 (.93)	7	3.02 (.57)
Openness	8	3.58 (.45)	6	3.22 (.31)	8	3.42 (.66)	7	3.02 (.75)
Agreeableness	8	3.94 (.58)	6	3.58 (.43)	8	3.90 (.53)	7	3.74 (.47)
Conscientiousness	8	3.81 (.73)	6	3.42 (.73)	8	3.96 (.69)	7	3.33 (.80)
<u>Rating 11</u>								
Domain	<i>n</i>	Client	<i>n</i>	Therapist	<i>n</i>	Client	<i>n</i>	Therapist
Neuroticism	5	3.07 (.69)	4	3.63 (.40)	3	2.61 (.68)	2	3.75 (.35)

Table 3 continued

Extraversion	5	2.97 (.34)	4	3.00 (.59)	3	3.33 (.44)	2	3.00 (.71)
Openness	5	3.33 (.46)	4	2.84 (.45)	3	3.06 (.20)	2	3.17 (1.41)
Agreeableness	5	3.57 (.77)	4	3.29 (.25)	3	3.89 (1.07)	2	3.75 (.35)
Conscientiousness	5	3.87 (.88)	4	3.54 (.63)	3	3.89 (.54)	2	3.25 (.12)
<u>Rating 13</u>								
Domain	<i>n</i>	Client	<i>n</i>	Therapist				
Neuroticism	2	3.00 (.71)	1	3.17				
Extraversion	2	2.92 (.59)	1	3.67				
Openness	2	3.34 (.23)	1	3.83				
Agreeableness	2	3.67 (.23)	1	4.00				
Conscientiousness	2	3.67 (.71)	1	3.17				

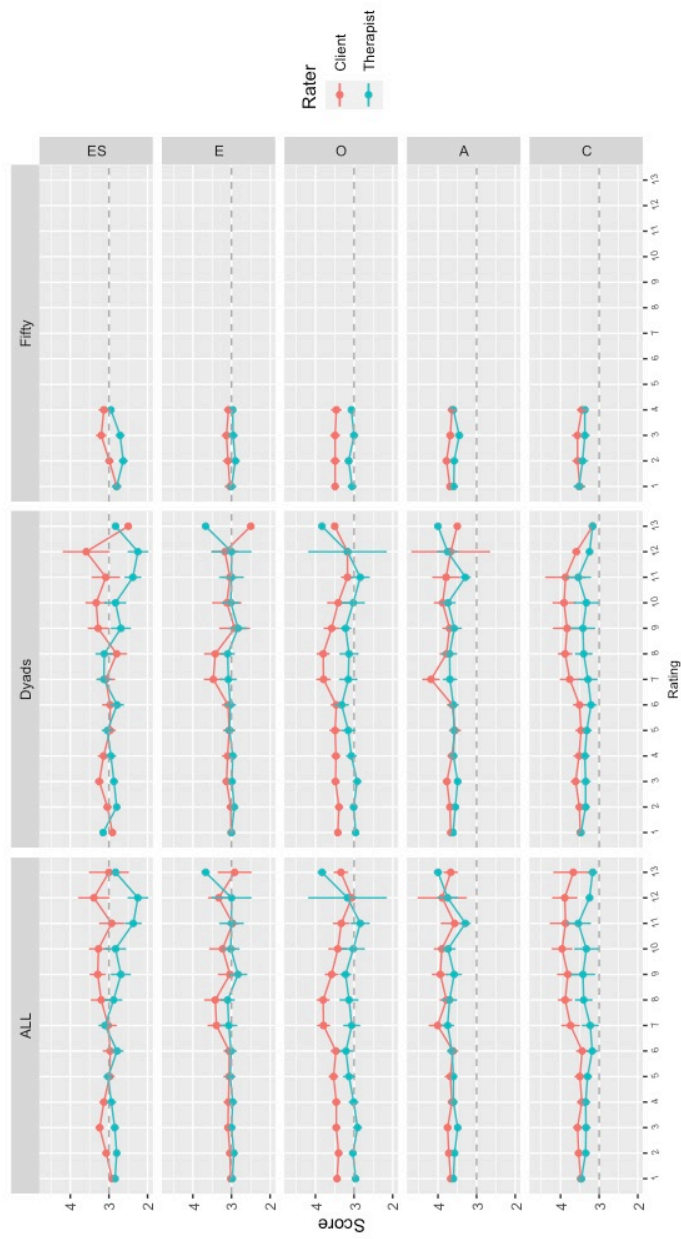


Figure 1. Mean-level personality trait trends for client and therapist ratings at each assessment.

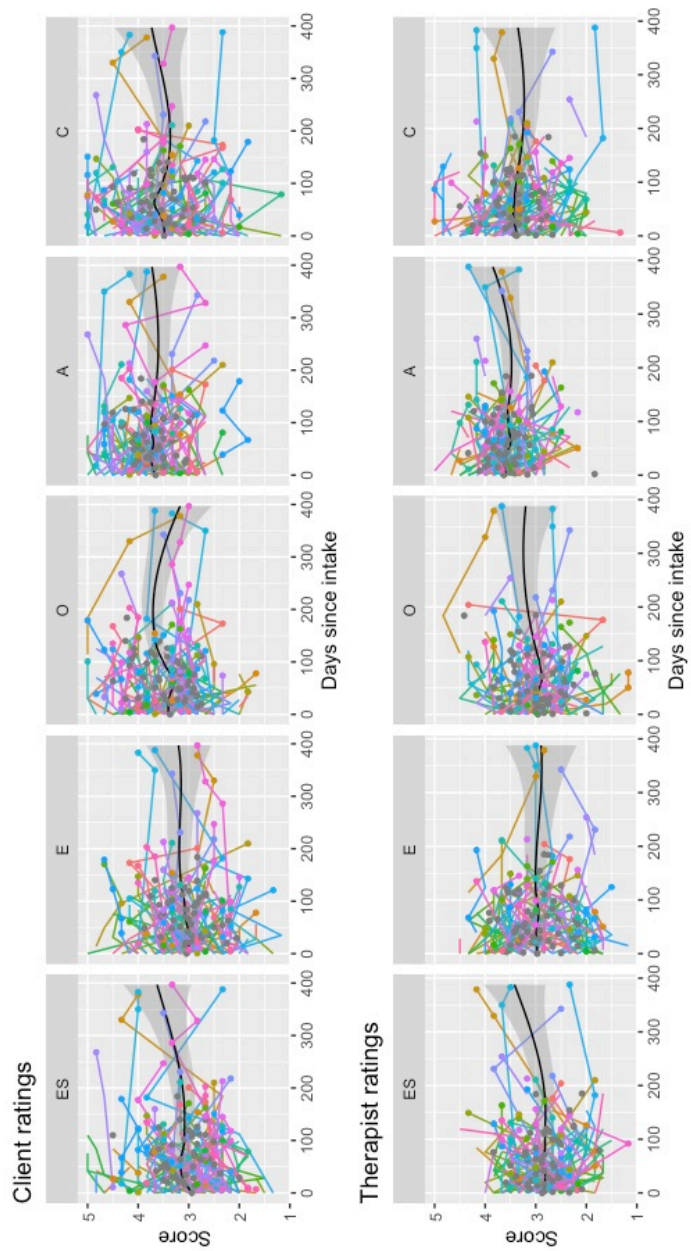


Figure 2. Spaghetti plot of all client and therapist ratings across days used in analyses.

Table 4. Domain Means and Standard Deviations for Each Rating for a Subset Sample

Domain	<i>n</i>	<u>Rating 1</u>		<i>n</i>	Therapist	<i>n</i>	<u>Rating 2</u>		<i>n</i>	Therapist
		Client					Client			
Neuroticism	50	3.20 (.71)		47	3.20 (.53)	45	3.01 (.72)		41	3.37 (.59)
Extraversion	50	3.04 (.67)		47	3.00 (.79)	45	3.09 (.73)		41	2.90 (.77)
Openness	50	3.49 (.69)		47	3.05 (.58)	45	3.49 (.65)		41	3.14 (.71)
Agreeableness	50	3.69 (.65)		47	3.59 (.42)	45	3.78 (.62)		41	3.58 (.51)
Conscientiousness	50	3.50 (.75)		47	3.52 (.79)	45	3.56 (.84)		41	3.43 (.76)
	<i>n</i>	<u>Rating 3</u>		<i>n</i>	Therapist	<i>n</i>	<u>Rating 4</u>		<i>n</i>	Therapist
		Client					Client			
Neuroticism	49	2.80 (.71)		41	3.29 (.75)	50	2.87 (.73)		40	3.05 (.74)
Extraversion	49	3.13 (.70)		41	2.96 (.68)	50	3.08 (.79)		40	2.97 (.65)
Openness	49	3.49 (.62)		41	3.00 (.72)	50	3.46 (.67)		40	3.07 (.74)
Agreeableness	49	3.68 (.65)		41	3.45 (.57)	50	3.64 (.63)		40	3.60 (.57)
Conscientiousness	49	3.56 (.92)		41	3.37 (.85)	50	3.45 (.81)		40	3.37 (.67)

Note. The subset consisted of 50 clients who completed the fourth rating. Five of these clients did not complete rating 2, and one did not complete rating 3.

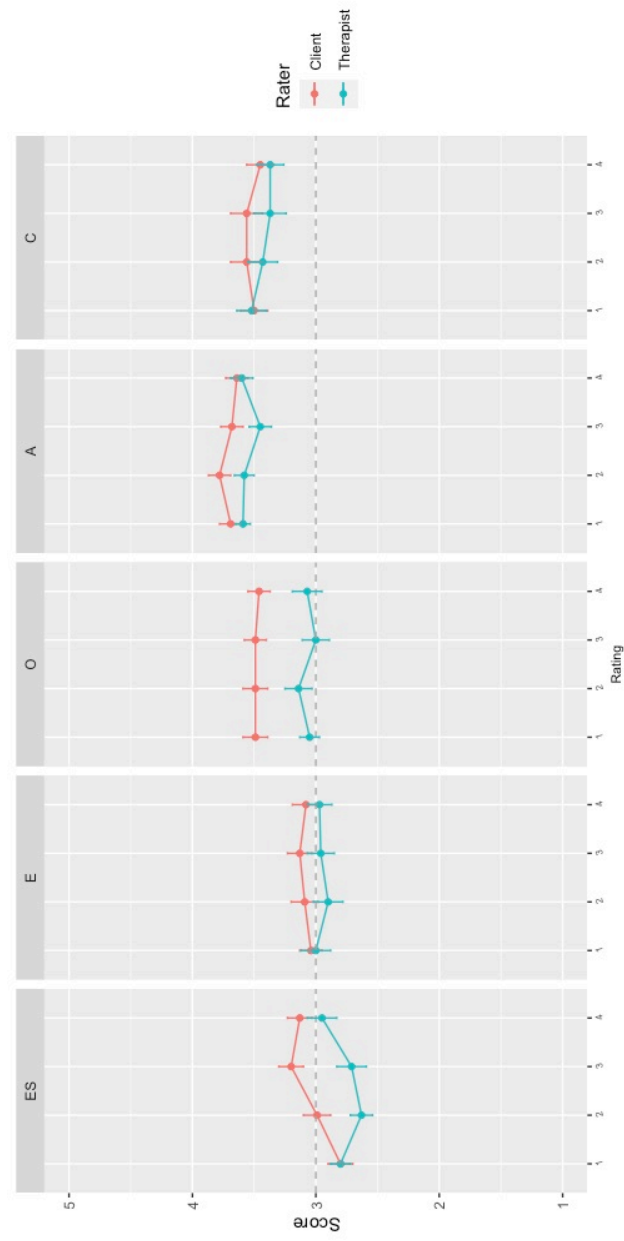


Figure 3. Mean-level personality trait trends for a subset ($n = 50$) of clients through the first four ratings.

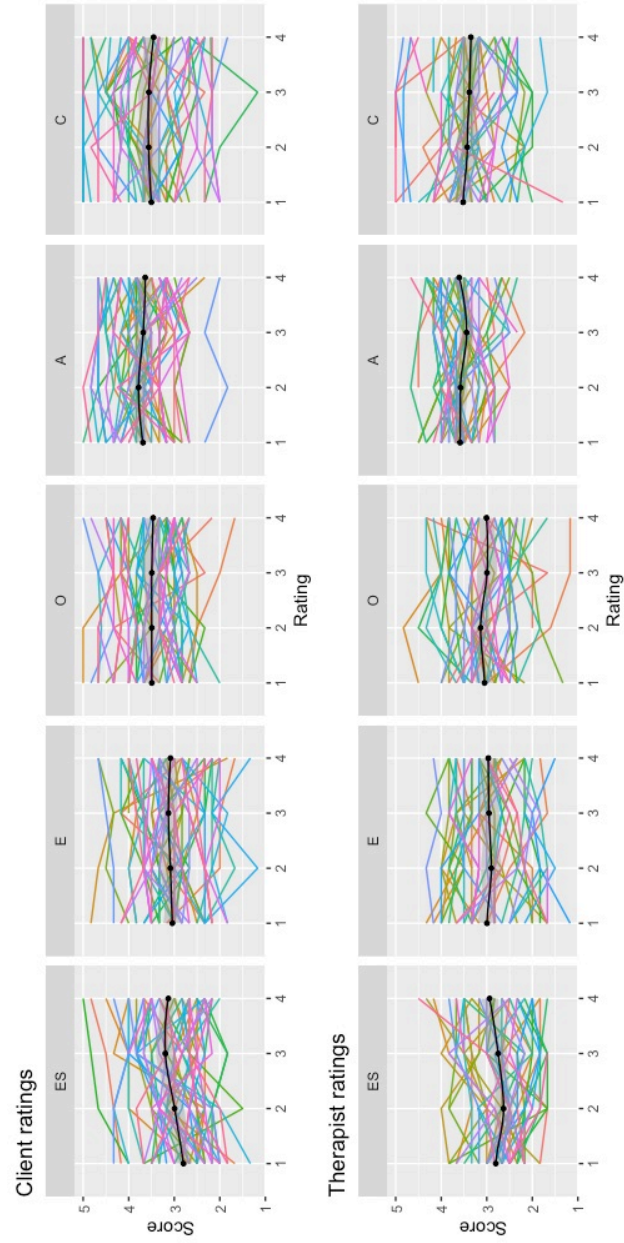


Figure 4. Spaghetti plots for a subset ($n = 50$) clients that at least completed up to the fourth rating.

Table 5. Linear Mixed-Effects Model of Client and Therapist Personality Ratings Over Time, not Including Distress

	Est.	SE	df	t
Client Ratings				
Emotional Stability	2.99	.07	197	44.63****
Time	.002	.00	197	3.75***
Extraversion	3.02	.07	198	46.32****
Time	.00	.00	198	.54
Openness	3.39	.06	198	56.00****
Time	.00	.00	198	.03
Agreeableness	3.69	.05	198	67.56****
Time	.00	.00	198	-.95
Conscientiousness	3.54	.07	197	49.97****
Time	.00	.00	197	-.92
Therapist Ratings				
Emotional Stability	2.83	.05	175	54.74****
Time	.00	.00	175	1.12
Extraversion	2.96	.07	175	43.01****
Time	.00	.00	175	-.01
Openness	2.96	.05	174	55.61
Time	.00	.00	174	-.92
Agreeableness	3.58	.05	175	75.32****
Time	.00	.00	175	-1.36
Conscientiousness	3.43	.07	175	50.11****
Time	-.002	.00	175	-2.96**

Note. * $p < .05$. ** $p < .01$. *** $p < .001$. **** $p < .000$.

Table 6. Linear Mixed-Effects Model of Client and Therapist Personality Ratings Over Time, Including Distress as a Covariate

	<i>Est.</i>	<i>SE</i>	<i>df</i>	<i>t</i>
Client Ratings				
Emotional Stability	3.75	.09	193	42.15****
Distress	-.60	.06	193	-10.47****
Time	.00	.00	193	1.24
Extraversion	3.49	.10	194	35.69****
Distress	-.36	.06	194	-6.08****
Time	.00	.00	194	-1.02
Openness	3.52	.10	194	36.46****
Distress	-.10	.06	194	-1.65
Time	.00	.00	194	-.45
Agreeableness	3.79	.09	194	42.73****
Distress	-.08	.05	194	-1.48
Time	.00	.00	194	-1.41
Conscientiousness	3.66	.11	193	33.12****
Distress	-.10	.07	193	-1.40
Time	.00	.00	193	-1.35
Therapist Ratings				
Emotional Stability	3.21	.09	169	35.70****
Distress	-.30	.06	169	-5.08****
Time	.00	.00	169	-.08
Extraversion	3.26	.11	169	30.42****
Distress	-.23	.06	169	-3.57***
Time	.00	.00	169	-.87
Openness	3.00	.10	168	31.49****
Distress	-.02	.06	168	-.48
Time	.00	.00	168	-.73

Table 6 continued

	<i>Est.</i>	<i>SE</i>	<i>df</i>	<i>t</i>
Agreeableness	3.68	.09	169	43.24****
Distress	-.08	.05	169	-1.53
Time	.00	.00	169	-1.70
Conscientiousness	3.39	.11	169	29.93****
Distress	.02	.07	169	.34
Time	-.001	.00	169	-2.64**

Note. Distress is measured via the CORE; Time is conceptualized as days from the intake session.

* $p < .05$. ** $p < .01$. *** $p < .001$. **** $p < .000$.

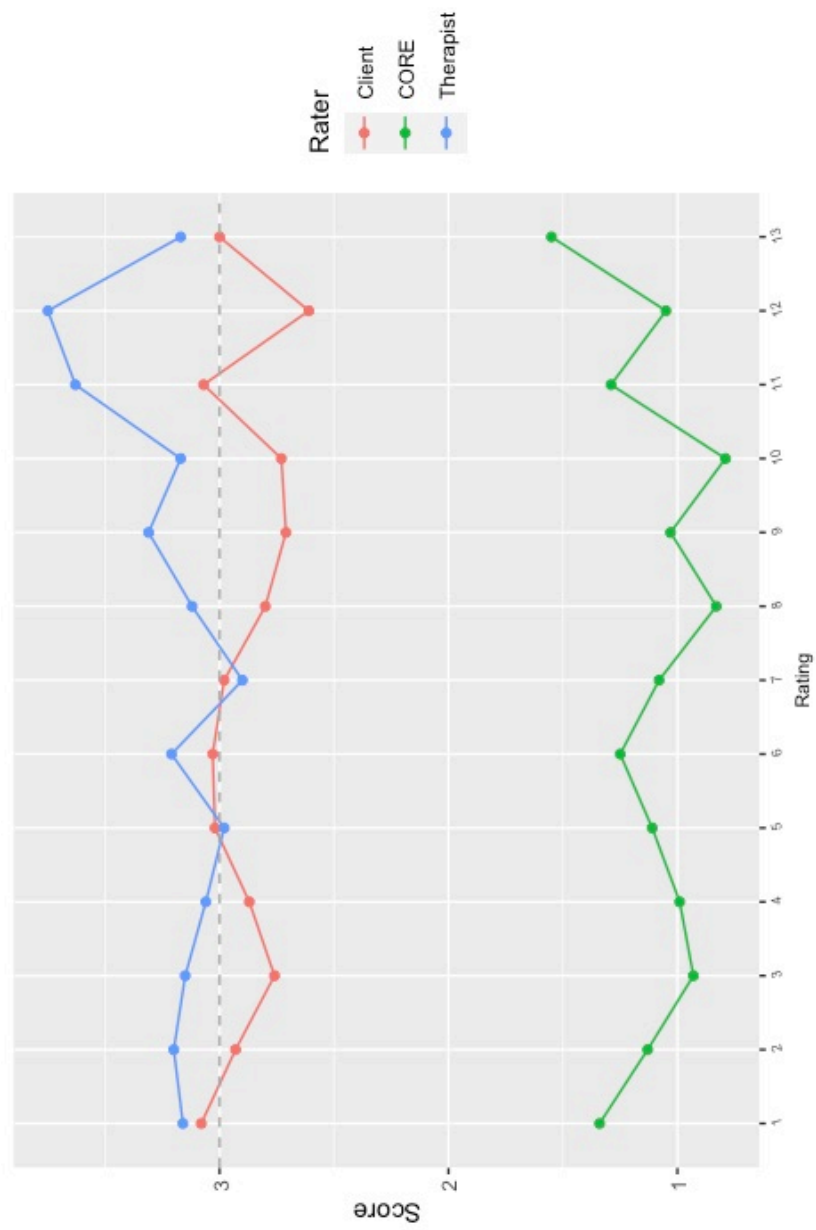


Figure 5. Therapist- and client-reported neuroticism plotted with client's ratings of distress throughout treatment.

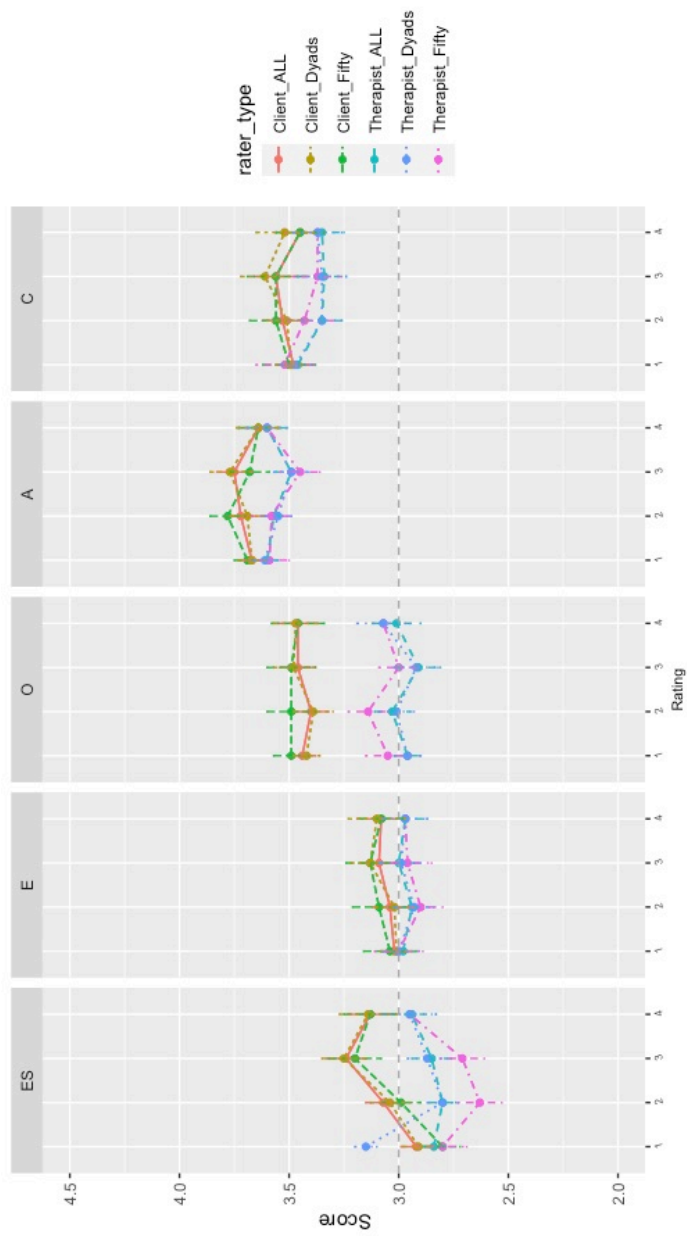


Figure 6. An overlay of the first four personality ratings for the entire sample, a subset of 50 with the first four ratings, and those with ratings at each time point that included both dyads.

Client-Therapist Agreement Across Sessions

Rank-order agreement across the first four ratings showed high agreement (median r s ranged from .40 for time 1 to .60 to time 4) across all five domains that were medium effects or larger (Table 7), when not controlling for distress. Similar patterns were generally seen in the partial correlations controlling for clients' self-reported distress (Table 8) except for neuroticism, which showed decreases in agreement over time, particularly for rating 4 ($r = .09$). Agreement between client and therapists' ratings of extraversion at rating 4 also decreased when controlling for clients' distress.

Paired t -tests (Table 9) showed that clients consistently rated themselves as higher in openness to experience compared to their therapists across the first four ratings (Cohen's d ranged from .53 to .76). Also of note, although client- and therapist-reported levels of neuroticism at the first rating were fairly similar with one another ($d = -.09$), their ratings started to diverge at rating 2, such that clients rated themselves as lower in neuroticism compared to their therapist ($d = -.36$). These trends continue throughout ratings 3 ($d = -.51$) and 4 ($d = -.25$). That is, clients and therapists reported similar levels of neuroticism at the outset of treatment. Yet, over the course of the next two ratings clients reported significantly lower levels of neuroticism compared to therapists, who reported similar levels from rating 1. A similar pattern was seen for the subset of 50 clients who completed the first four assessments (Table 4), suggesting these differences are not attributable to only attrition. A similar pattern was also seen for agreeableness, such that client and therapist ratings were quite similar during the first rating, but the therapist rated the client as lower in agreeableness at rating 2 ($d = .25$) and 3 ($d = .45$). Similar to what was reported in the linear model, therapists also rated the client as lower in conscientiousness after the first session, and this was most notable at rating 3 ($d = .31$).

Absolute agreement (assessed via ICCs) between all client-therapist dyads was also examined across time (Table 10; Figure 7). Similar to before, average scores of a subset of 39 dyads that completed up to rating 4 were also examined in order to identify whether trends in agreement might be due to attrition or actual trends in the sample (Table 11). In the entire sample, average ICCs of the domains ranged from .32 (conscientiousness) to .60 (extraversion), and the full-measure ICCs across time points averaged around .67. To further probe the trends in absolute agreement between clients and therapists over time, a second mixed-effects model was conducted that examined agreement across the entire FFMRF (Table 12). None of the variables were

significant predictors in the model indicating that, first, agreement did not linearly increase or decrease over time. Additionally, the number of days in between ratings or whether the client switched therapists did not significantly affect agreement. Lastly, the clients' negative impressions of themselves did not impact agreement.

Table 7. Correlations Between Therapist and Client Ratings

Time 1 (<i>n</i> = 121)										Time 2 (<i>n</i> = 72)									
Therapist Ratings					Client Ratings					Therapist Ratings					Client Ratings				
	N	E	O	A	C		N	E	O	A	C		N	E	O	A	C		
N	.45	-.36	-.02	-.07	-.06		N						.42	-.18	-.05	-.01	.00		
E	-.23	.40	.19	-.25	.00		E						-.22	.58	.10	-.24	-.10		
O	.19	-.07	.38	-.06	-.19		O						.16	.01	.46	-.22	-.43		
A	-.04	.03	.07	.32	.17		A						-.04	-.03	.22	.33	.01		
C	-.12	.02	-.06	.23	.48		C						-.09	-.11	-.20	.22	.47		
Time 3 (<i>n</i> = 57-58)										Time 4 (<i>n</i> = 40)									
Therapist Ratings					Client Ratings					Therapist Ratings					Client Ratings				
	N	E	O	A	C		N	E	O	A	C		N	E	O	A	C		
N	.44	-.23	.01	-.27	-.25		N						.32	-.30	.00	-.34	-.40		
E	-.19	.59	.07	.00	-.03		E						-.19	.35	.17	-.01	.06		
O	.08	-.01	.61	.04	-.26		O						-.02	.25	.62	.01	-.19		
A	-.16	-.09	-.03	.50	.23		A						-.02	.26	.17	.60	.14		
C	-.24	.22	-.26	.20	.68		C						-.11	.08	-.23	.23	.68		

Note. The letter corresponds with the specific domain, and the number corresponds with the assessment time point. The third rating of openness to experience only had 57 dyads, whereas the other four domains had 58 dyads.

Table 8. Partial Correlations Between Therapist and Client Ratings, Controlling for Distress

Time 1										Time 2									
Therapist Ratings					Client Ratings (<i>n</i> = 115)					Therapist Ratings					Client Ratings (<i>n</i> = 67)				
	N	E	O	A	C		N	E	O	A	C		N	E	O	A	C		
N	-.36	-.27	.02	-.05	-.06		N						.25	-.02	-.05	.04	.05		
E	-.10	.36	.14	-.26	.00		E						-.01	.53	.07	-.29	-.15		
O	.31	-.09	.32	-.02	-.21		O						.06	.04	.39	-.15	-.44		
A	-.01	-.01	.05	.30	.18		A						-.06	-.01	.28	.31	.01		
C	-.22	.06	-.06	.24	.51		C						-.08	-.14	-.20	.20	.47		
Time 3										Time 4									
Therapist Ratings					Client Ratings (<i>n</i> = 53)					Therapist Ratings					Client Ratings (<i>n</i> = 37)				
	N	E	O	A	C		N	E	O	A	C		N	E	O	A	C		
N	.26	-.09	.00	-.17	-.17		N						.09	-.12	-.01	-.26	-.34		
E	.04	.52	.02	-.07	-.09		E						.02	.23	.20	-.11	-.02		
O	.16	-.08	.58	.09	-.26		O						.07	.23	.63	-.03	-.23		
A	.01	-.21	-.01	.45	.16		A						.02	.27	.17	.61	.13		
C	-.15	.18	-.26	.13	.67		C						-.11	.08	-.23	.23	.68		

Note. The letter corresponds with the specific domain, and the number corresponds with the assessment time point. The relationship at each time point is controlled for by client's report of distress (measured via the CORE).

Table 9. Paired Samples T-tests of Client and Therapist Ratings

<u>Rating 1</u>						
Domain	<i>n</i>	Client	Therapist	<i>t</i>	<i>p</i>	<i>CI</i> <i>d</i>
Neuroticism	121	3.09 (.76)	3.15 (.59)	-.79	.43	-.18, .08
Extraversion	121	3.01 (.72)	3.00 (.78)	.22	.83	-.13, .17
Openness	121	3.42 (.71)	2.96 (.61)	6.81	.00	.33, .60
Agreeableness	121	3.67 (.63)	3.62 (.54)	.91	.36	-.07, .18
Conscientiousness	121	3.49 (.82)	3.47 (.78)	.26	.80	-.13, .16
<u>Rating 2</u>						
Domain	<i>n</i>	Client	Therapist	<i>t</i>	<i>p</i>	<i>CI</i> <i>d</i>
Neuroticism	72	2.96 (.76)	3.20 (.56)	-2.74	.01	-.41, -.06
Extraversion	72	3.02 (.73)	2.93 (.74)	1.21	.23	-.06, .25
Openness	72	3.39 (.76)	3.01 (.67)	4.27	.00	.20, .55
Agreeableness	72	3.69 (.60)	3.55 (.52)	1.84	.07	-.01, .29
Conscientiousness	72	3.51 (.79)	3.35 (.78)	1.70	.09	-.03, .35
<u>Rating 3</u>						
Domain	<i>n</i>	Client	Therapist	<i>t</i>	<i>p</i>	<i>CI</i> <i>d</i>
Neuroticism	58	2.75 (.78)	3.13 (.71)	-3.68	.00	-.59, -.17
Extraversion	58	3.13 (.66)	2.99 (.68)	1.74	.09	.08, -.02
Openness	57	3.47 (.69)	2.92 (.75)	6.62	.00	.39, .72

Table 9 continues

Agreeableness	58	3.77 (.66)	3.49 (.58)	3.48	.00	.12, .45	.45	
Conscientiousness	58	3.61 (.86)	3.35 (.82)	2.98	.00	.09, .44	.31	
<u>Rating 4</u>								
	<i>n</i>	Client	Therapist	<i>t</i>	<i>p</i>	<i>CI</i>	<i>d</i>	
Neuroticism	40	2.86 (.80)	3.05 (.75)	-1.37	.18	-.48, .09	-.25	
Extraversion	40	3.10 (.79)	2.97 (.65)	1.05	.30	-.13, .40	.18	
Openness	40	3.47 (.64)	3.07 (.74)	4.21	.00	.21, .60	.58	
Agreeableness	40	3.64 (.64)	3.60 (.57)	.48	.64	-.13, .22	.07	
Conscientiousness	40	3.52 (.83)	3.37 (.67)	1.55	.13	-.05, .35	.20	

Note. *n* = number of pairs in analyses. Multiply this number by two for the number of all participants included in each analysis.
CI = Confidence intervals for mean differences. *d* = Cohen's *d* effect sizes.

Table 10. Average ICCs for Each Rating

	Rating 1	Rating 2	Rating 3	Rating 4
Domain	<i>n</i> = 121	<i>n</i> = 72	<i>n</i> = 57-58	<i>n</i> = 40
Neuroticism	.63	.64	.48	.56
Extraversion	.59	.65	.61	.55
Openness	.47	.45	.50	.49
Agreeableness	.40	.46	.43	.44
Conscientiousness	.31	.30	.37	.30
Entire FFMRF	.66	.71	.68	.63

Note. On rating 3, conscientiousness had one less ICC coefficient than the other four domains.

Table 11. Average ICCs for Thirty Client-Therapist Dyads During the First Four Ratings

	Rating 1	Rating 2	Rating 3	Rating 4
Domain	<i>n</i> = 37	<i>n</i> = 32	<i>n</i> = 36	<i>n</i> = 39
Neuroticism	.63	.65	.44	.56
Extraversion	.61	.67	.59	.55
Openness	.48	.48	.48	.48
Agreeableness	.44	.41	.45	.46
Conscientiousness	.30	.38	.33	.30
Entire FFMRF	.71	.72	.67	.63

Note. The ICCs were a subset of client-therapist dyads that had completed the fourth rating. Two dyads did not have agreement coefficients for rating 1, seven did not have an agreement coefficient for rating 2, and three did not have a coefficient for rating three.

Table 12. Linear Mixed-Effects Model of Client-Therapist Agreement

	<i>Est.</i>	<i>SE</i>	<i>df</i>	<i>t</i>
Intercept	.72	.03	112	26.49****
Time	.00	.00	112	-.88
Days Between Raters	.00	.00	112	.19
Negative Impression Management	.00	.00	71	1.42
Therapist Change	.02	.08	71	.19

Note. * $p < .05$. ** $p < .01$. *** $p < .001$. **** $p < .0001$.

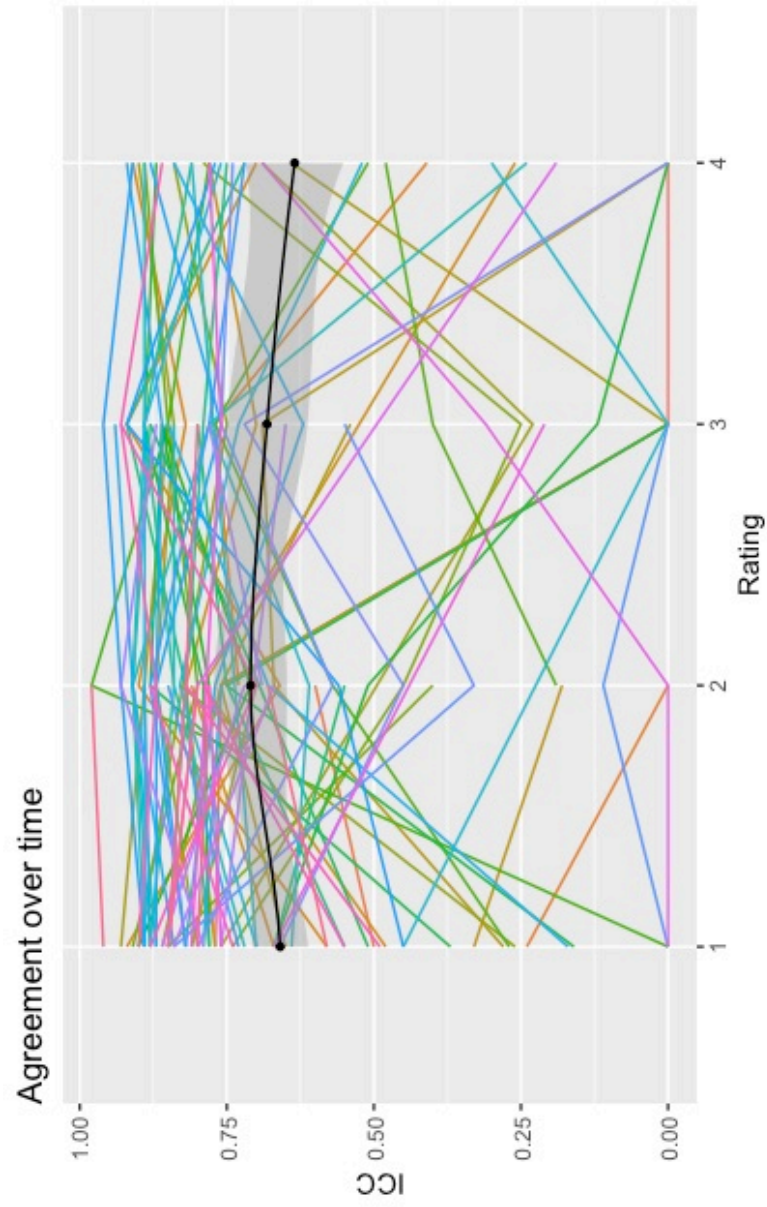


Figure 7. Plotted agreement (ICCs) across the first four assessment time points.

Agreement as Predictors of Engagement and Improvement

As can be seen in Table 1, the average number of sessions canceled, rescheduled, or no-showed by the clients were 3.85, or 21% of sessions scheduled. Average RCI was -.68, with a wide range of change in both directions across the sample. Thirteen clients had reliable decreases in symptoms (i.e., had scores at or lower than -1.96; an additional three more clients had scores lower than -1.90 but did not quite meet the reliable threshold). One client had reliable increase in reported symptoms (i.e., above 1.96).

All variables in each regression analysis were correlated with each other (Table 13). The only noteworthy correlation between predictor and outcome variables were engagement's relation to Time 1 agreement and attendance's relation to Time 2 agreement, overall small effects ($r = .21$). In other words, the more the client and therapist agreed at Time 1, the more engaged the client was in treatment (i.e., the more sessions that were scheduled and attended). Table 13 also indicated a pattern of relationships among agreement at subsequent time points that ranged from generally medium to large effects, such that the more the client and therapist agreed at one session, the more likely they were to agree at subsequent sessions. When examining the outcome variables' relations with one another, it is worth noting that the variable "engagement" was essentially identical to the number of sessions attended. Thus, it is likely unnecessary to include both of these in the analyses; however, all analyses are included in this document for full transparency.

Across all regression analyses conducted (Tables 14 and 15), R^2 s ranged from .04 to .20, none of which were statistically significant. None of the predictors significantly predicted the outcome variables at $p < .01$. Rating 2 agreement did predict improvement at $p < .05$, such that lower levels of agreement were related to less improvement (i.e., increases in scores reported on the CORE outcome measure). When examining attendance and engagement, agreement at rating 4 predicted number of sessions attended and total engagement ($p < .05$). That is, lower levels of agreement at the fourth rating predicted more sessions attended and more overall engagement.

Table 13. Correlations Among Improvement, Engagement, and Client-Therapist Agreement

	Engagement	% Attended	# Sessions	Time 1 Agreement	Time 2 Agreement	Time 3 Agreement	Time 4 Agreement
Improvement	-.05	-.15	-.08	-.16	-.08	.15	.10
Engagement	—	.10	.97	.21	.07	.10	-.06
% Attended	.10	—	.30	-.02	.22	.10	.00
# Sessions	.97	.30	—	.19	.16	.09	-.07
Time 1 Agreement	.21	-.02	.19	—	.51	.38	.26
Time 2 Agreement	.07	.22	.16	.51	—	.55	.28
Time 3 Agreement	.10	.10	.09	.38	.55	—	.50
Time 4 Agreement	-.06	.00	-.07	.26	.28	.50	—

Note. Bolded coefficients indicate small or larger effects. Domain scores indicate absolute agreement, or ICCs, among clients and therapists. Improvement = Reliable Change Index score. Engagement = weighted engagement score consisting of percentage of sessions attended weighted by number of sessions attended. % Attended = percentage of sessions attended. # Sessions = number of sessions attended.

Table 14. Linear Regression of Agreement Predicting Improvement

	<i>B</i>	95% CI	<i>SE</i>
Rating	$R^2 = .20, F = 2.65$		
1	-.64	-2.72, 1.44	1.00
2	-3.25*	-6.13, -.38	1.39
3	1.99	-.07, 4.05	1.00
4	.18	-1.59, 1.96	.86

Note. Improvement was calculated as the RCI on the CORE. Agreement is calculated by ICCs.

* $p < .05$.

Table 15. Linear Regressions of Agreement Predicting Attendance and Engagement

	<i>B</i>	95% CI	<i>SE</i>
Percentage of sessions attended		$R^2 = .04, F = 1.26$	
Rating 1	-.10	-.39, .19	.14
Rating 2	.20	-.14, .55	.16
Rating 3	.10	-.09, .30	.09
Rating 4	-.05	-.21, .12	.08
Number of sessions attended		$R^2 = .07, F = 1.47$	
Rating 1	11.54	-9.29, 32.36	10.07
Rating 2	1.05	-27.75, 29.85	13.92
Rating 3	4.35	-16.29, 24.99	9.98
Rating 4	-18.68*	-36.45, -.91	.04
Total Engagement		$R^2 = .14, F = 2.03$	
Rating 1	23.65	-3.06, 50.35	12.84
Rating 2	-12.69	-44.03, 18.66	15.07
Rating 3	6.31	-11.39, 24.00	8.51
Rating 4	-17.90*	-33.21, -2.60	7.36

Note. Ratings were run separately from one another. Engagement was quantified as percentage of sessions attended weighted with the number.

* $p < .05$.

Supplementary, Exploratory, and Post-hoc Analyses

As proposed, I compared the facet-level mean scores for the first four ratings for clients and therapists in an exploratory fashion (Table 16). Client-therapist ratings were compared with Cohen's d effect sizes to identify the overall effect in mean-level differences. The largest effects were generally seen for facets related to openness to experience such that, throughout treatment, clients consistently rated themselves as higher in openness to feelings, ideas, and values. They also consistently rated themselves as higher on tender-mindedness. While not seen in the first rating, clients tended to report themselves as lower in anxiousness, angry hostility, self-consciousness, and vulnerability compared to their therapists in subsequent ratings.

Additional post-hoc analyses that were not proposed were further conducted to examine change in mean domain-level ICCs across the first four ratings of each specific domain. ICC comparisons were conducted using bootstrapping resampling of the difference between the mean ICCs of two time points. Ten thousand samples were conducted and based on the mean difference in each of these samples, 95% confidence intervals were estimated. Across all five domains, neuroticism was the only domain that had confidence intervals that were statistically significant. There was no significant difference between the first and second rating, but ratings one and two were both significantly higher than ratings three and four, suggesting there were noticeable decreases in absolute agreement from the outset of treatment.

Table 16. Facet-Level Scores per Rating for Clients and Therapists.

Facet	<u>Rating 1</u>		<u>Rating 2</u>		<u>Rating 3</u>		<u>Rating 4</u>	
	Client (n = 126)	Therapist (n = 123)	Client (n = 84)	Therapist (n = 79)	Client (n = 70)	Therapist (n = 60)	Client (n = 50)	Therapist (n = 43)
N1 Anxiousness	3.69 (1.10)	3.72 (.97)	3.40 (1.14)	3.66 (.97)	3.01 (1.01)	3.53 (.96)	3.46 (1.05)	3.21 (.94)
N2 Angry Hostility	2.33 (1.19)	2.37 (1.03)	2.21 (1.11)	2.53 (1.07)	2.25 (1.21)	2.65 (1.07)	2.38 (1.03)	2.60 (1.03)
N3 Depressiveness	3.32 (1.19)	3.25 (1.09)	3.01 (1.18)	3.18 (1.05)	2.83 (1.14)	3.05 (1.00)	2.86 (1.23)	2.95 (.95)
N4 Self-consciousness	3.38 (1.23)	3.52 (1.09)	3.33 (1.27)	3.62 (.96)	3.06 (1.09)	3.42 (1.03)	3.20 (1.16)	3.33 (1.04)
N5 Impulsivity	2.80 (1.24)	2.74 (1.11)	2.81 (1.26)	2.89 (1.15)	2.71 (1.18)	2.93 (1.09)	2.50 (1.09)	2.88 (1.07)
N6 Vulnerability	2.98 (1.18)	3.35 (1.00)	2.81 (1.09)	3.33 (1.03)	2.74 (.96)	3.33 (1.00)	2.76 (1.13)	3.40 (1.03)
E1 Warmth	3.73 (1.02)	3.57 (1.01)	3.72 (1.08)	3.57 (1.07)	3.68 (1.04)	3.67 (.95)	3.56 (1.28)	3.44 (.93)
E2 Gregariousness	3.09 (1.16)	3.15 (1.19)	3.19 (1.06)	3.04 (1.15)	3.20 (1.07)	3.15 (.99)	3.14 (1.16)	3.14 (1.06)
E3 Assertiveness	2.74 (1.11)	2.63 (1.05)	2.58 (1.23)	2.61 (1.11)	2.76 (1.04)	2.75 (.93)	2.86 (1.16)	2.65 (.84)
E4 Activity	2.92 (1.11)	3.08 (1.15)	3.04 (1.04)	3.04 (1.08)	3.01 (1.13)	2.97 (1.07)	3.12 (1.21)	2.93 (.99)
E5 Excitement-seeking	2.65 (1.14)	2.57 (.94)	2.55 (1.11)	2.63 (1.00)	2.52 (1.04)	2.57 (.87)	2.56 (1.01)	2.58 (.85)
E6 Positive emotions	2.96 (1.11)	2.84 (1.15)	3.21 (1.13)	2.77 (1.07)	3.26 (.95)	2.88 (.88)	3.26 (1.10)	3.05 (.90)
O1 Fantasy	3.21 (1.25)	2.68 (.99)	3.19 (1.41)	2.96 (1.15)	3.26 (1.30)	2.71 (1.08)	3.32 (1.28)	2.81 (1.16)
O2 Aesthetics	3.43 (1.08)	3.05 (.84)	3.49 (1.05)	3.14 (.85)	3.35 (.95)	3.10 (.92)	3.46 (1.01)	3.23 (.95)
O3 Feelings	3.90 (1.00)	3.59 (1.02)	4.07 (.78)	3.41 (1.07)	3.91 (.95)	3.29 (1.10)	4.00 (.95)	3.30 (1.04)
O4 Actions	2.88 (1.01)	2.66 (.93)	2.65 (1.12)	2.63 (.96)	2.88 (1.07)	2.56 (.95)	2.80 (1.16)	2.77 (.95)
O5 Ideas	3.50 (1.06)	2.74 (.87)	3.33 (1.22)	2.92 (.86)	3.51 (1.03)	2.66 (.92)	3.48 (.95)	2.79 (1.04)
O6 Values	3.70 (1.12)	3.00 (.88)	3.66 (1.06)	3.08 (.90)	3.84 (.92)	3.17 (.97)	3.66 (1.08)	3.12 (.96)
A1 Trust	3.02 (1.21)	3.04 (1.01)	3.02 (1.10)	3.10 (1.02)	3.04 (1.08)	2.98 (1.02)	3.10 (1.22)	3.14 (.99)
A2 Straightforwardness	3.87 (1.05)	3.94 (.74)	3.98 (.94)	3.77 (.80)	3.84 (.95)	3.58 (.81)	3.82 (.92)	3.81 (.82)
A3 Altruism	3.89 (.97)	3.61 (.77)	3.81 (.94)	3.63 (.77)	3.93 (.96)	3.60 (.79)	3.88 (.92)	3.70 (.80)

Table 16 continued

Facet	<u>Rating 1</u>		<u>Rating 2</u>		<u>Rating 3</u>		<u>Rating 4</u>	
	Client (<i>n</i> = 126)	Therapist (<i>n</i> = 123)	Client (<i>n</i> = 84)	Therapist (<i>n</i> = 79)	Client (<i>n</i> = 70)	Therapist (<i>n</i> = 60)	Client (<i>n</i> = 50)	Therapist (<i>n</i> = 43)
A4 Compliance	3.60 (<i>1.03</i>)	3.90 (<i>.81</i>)	3.69 (1.01)	3.90 (.91)	3.84 (1.01)	3.72 (.87)	3.44 (.88)	3.67 (.81)
A5 Modesty	3.66 (1.06)	3.53 (.91)	3.69 (.97)	3.49 (1.07)	3.77 (.99)	3.73 (1.01)	3.74 (.96)	3.56 (1.01)
A6 Tender-mindedness	3.98 (.90)	3.62 (.86)	4.07 (.85)	3.51 (.80)	4.04 (.81)	3.45 (.81)	3.84 (.82)	3.60 (.82)
C1 Competence	3.58 (1.03)	3.41 (1.03)	3.52 (1.04)	3.41 (1.02)	3.57 (.96)	3.20 (<i>1.12</i>)	3.60 (<i>1.12</i>)	3.33 (.89)
C2 Order	3.23 (1.18)	3.20 (1.14)	3.24 (1.20)	3.14 (1.02)	<i>3.43 (1.10)</i>	<i>3.18 (1.08)</i>	3.18 (1.22)	3.16 (1.00)
C3 Dutifulness	3.70 (1.06)	3.54 (.89)	3.69 (<i>1.08</i>)	3.47 (.93)	3.62 (1.09)	3.63 (.96)	3.62 (1.07)	3.63 (.87)
C4 Achievement	3.49 (1.14)	3.50 (.93)	3.57 (<i>1.10</i>)	3.30 (.90)	3.52 (1.16)	3.35 (1.04)	3.39 (1.04)	3.26 (.88)
C5 Self-discipline	3.26 (<i>1.12</i>)	3.47 (.89)	<i>3.42 (1.14)</i>	3.12 (.88)	3.39 (1.10)	3.23 (.91)	3.34 (1.02)	3.33 (.81)
C6 Deliberation	3.68 (1.03)	3.62 (1.02)	3.72 (.98)	3.58 (1.02)	<i>3.77 (1.06)</i>	<i>3.48 (.95)</i>	3.59 (1.00)	3.47 (.93)

Note. Italicized numbers indicate small effects (Cohen’s *d*); bold indicates medium-to-large effects.

DISCUSSION

The present study investigated the longitudinal trajectory of client personality traits as rated by both the client and therapist. Having ratings from both the client and therapist allowed for the examination of mean-level, rank-order, and absolute agreement between clients and therapist at not only the outset and end of treatment, but also throughout the intervention, and this was one of the first studies to be able to specifically do this. Additionally, this was one of the first studies with the ability to control for changes in state-level distress over time when accounting for personality trait change. The findings, which are discussed in more detail below, also highlight the importance of utilizing multi-method approaches for a more holistic understanding of complex questions.

Personality Change Across Time

The 207 studies analyzed in Roberts and colleagues (2017)'s meta-analysis generally showed moderate changes across all traits ($d = .37$), particularly in relation to increases on emotional stability and extraversion. However, as previously stated, a caveat to these findings were that the studies did not control for state-level distress, meaning they were unable to definitively state whether these changes were true trait change or due to changes in clients' states (i.e., distress). This study was able to control for clients' reported state-levels of distress when examining trait-level change in the hierarchical model. Findings suggested that there was a relationship with distress and personality ratings, such that higher levels of distress resulted in significantly lower levels of emotional stability and extraversion. This suggests that state-level change would likely be related to any trait-level change seen. In fact, when not controlling for distress, clients *did* report significant decreases in self-reported neuroticism, although this was not seen for therapists or for clients when controlling for distress.

The relationship between state- and trait-level change has been a topic of interest in the treatment literature for at least thirty years due to concerns that trait-level changes in therapeutic and pharmacological interventions might actually be due to state-level change (state-artifact position). Some have argued that traits and states cannot be separated from one another, and emotional stability/neuroticism is particularly influenced by state changes of distress (Griens, Jonker, Spinhoven, & Blom, 2002). Others have argued for the opposite – that changes in trait

influences the changes seen in one's state (cause-correction hypothesis; Soskin et al., 2012). In many other cases, studies have argued that state change was entirely or partially unrelated to trait change (Costa, Bagby, Herbst, & McCrae, 2005; Lorranger et al., 1991; Mullen, Blanco, Vaughan, Vaughan, & Roose, 1999).

The present study cannot clarify the specific relationship between clients' states and traits, but it is important to highlight the theorized bidirectional relationship between these two levels – that is, someone high in general trait neuroticism will likely also report high levels of state-level distress and being frequently in a state of distress will also likely result in a more habitual, prolonged pattern of behavioral distress representative of high neuroticism (Allemand & Flückiger, 2017; Rosenberg, 1998; Wrzus & Roberts, 2016). Thus, while high levels of distress were related to lower levels of adaptive personality functioning in this study, this likely is not a simple one-way relationship, and clients' personality traits might also have influenced their propensity to experience higher or lower levels of distress each session. It is also worth noting that although clients' initial decreases in self-reported neuroticism might be better reflected by their state-level distress decreasing, if they exhibit sustained low levels of distress, this could in theory reflect actual trait-level neuroticism decreasing (e.g., Allemand & Flückiger, 2017; Wrzus & Roberts, 2016). Thus, future research should continue to examine whether those who report lower levels of neuroticism exhibit sustained levels across time, both throughout treatment and after, in order to investigate true trait-level change.

Of note, neither clients nor therapists reported significant changes in extraversion, openness, agreeableness, and conscientiousness. Overall RCI scores of the CORE indicated that, as a whole, this sample generally decreased in their overall distress. If all traits were being influenced by the clients' state-level distress, increases in these scores would likely have been seen when decreases in distress were reported. Thus, it is likely that at least scores on extraversion, openness, agreeableness, and conscientiousness are representative of actual average trait stability. These traits were also likely not directly targeted in interventions like neuroticism might have been, which could also explain the limited changes seen in client and therapist report of these traits. Additionally, it might have been that some clients were aiming to change traits in different patterns. For example, while one client might have sought to increase maladaptively low levels of agreeableness, another client might have had goals more focused on decreasing maladaptively high levels of agreeableness. Future research should examine these more nuanced trait changes.

Another important consideration when examining mean-level changes in personality traits is the clients' baseline scores at the outset of treatment. Recently, Bleidorn and colleagues (2019) surveyed personality trait experts regarding their perception of the most adaptive personality profile on the FFMRF, the same measure used in the current study. Compared to the averages presented in their study, the present clients tended to rate themselves as generally lower in extraversion and higher in neuroticism, openness, agreeableness, and conscientiousness. Therapists also tended to rate their clients as lower in extraversion and higher in neuroticism and agreeableness. Because clients, on average, came into treatment already seeing themselves as incredibly adaptive on openness, agreeableness, and conscientiousness, these likely were not a part of treatment goals. Relatedly, for both these ratings as well as therapists' high ratings on agreeableness, it is probably unlikely to see significant increases in already high, adaptive levels of personality functioning. This highlights the importance of future replication, including studies examining trait change in clients with maladaptively low or high baseline levels of traits, as larger increases could potentially be seen.

This study was taken from a naturalistic sample in which interventions were not generally created to specifically target personality traits. It would likely be expected that the magnitude of trait change would be larger for those who are involved in specific interventions that target difficulties related to personality traits. Allan and colleagues (2018), for example, found quite large effects in personality trait change in a 10-week coaching program that was specifically designed to target change in personality traits, and this was particularly the case for facets related to neuroticism. Hudson and Fraley (2015) also found increases towards more desirable levels of personality functioning in an intervention designed to help participants attain their personality-change goals. Both naturalistic and experimental studies provide valuable information to the field as the combination of these studies could potentially inform therapists how to translate experimental studies to the real-world therapy room in which therapists might also need to focus on more high-needs treatment goals, such as distress tolerance and introducing coping skills to reduce risk of self-harm.

Client-Therapist Agreement Across Time

Few studies to date have investigated the relationship between client and therapist ratings of personality, particularly across time, and the majority of these studies have typically looked at

only one point in time. In the current sample, the median agreement across domains in the first rating (i.e., typically the intake session) was .40. This agreement is similar to that found in Connelly and Ones's (2010) meta-analytic report of agreement between family (*mdn* $r = .37$) and friends (*mdn* $r = .38$). Similarly, Oltmanns and Oltmanns (2021) found that, across all informants, the median meta-analytic correlation for personality pathology (as measured by the PID-5) was .45. They included ten studies that specifically looked at agreement between client and therapist ratings of personality pathology not just limited to the PID-5 which showed a median r of .25. Thus, although the current study did not use a measure specifically designed to assess only maladaptive personality functioning, the rank-order agreement seen throughout treatment in this sample was generally higher than previously reported.

Taken together, the findings highlight that the relationship between a client and therapist is unique. It is likely that due to the nature of the intake session in which clients comprehensively disclose personal information after having only known the therapist for less than 24 hours, therapists are able to develop a good general sense of the clients' personality functioning quickly. These median correlations only become stronger at subsequent sessions (i.e., time 2 = .46, time 3 = .59, time 4 = .53), suggesting that as the client and therapist progress throughout therapy, the rank-order agreement also improves. A similar pattern was seen when controlling for distress, such that median correlations went from moderate to large (i.e., time 2 = .39, time 3 = .52, time 4 = .61). An exception, however, was seen for neuroticism, such that, when controlling for distress, client and therapists' rank-order agreement became worse over time, with the correlation at time 4 being negligible. Again, this finding highlights how state change can be intertwined – or mistaken for – trait-level change and the importance of considering distress when examining clients' self-report of neuroticism in particular.

Although rank-order investigations can provide valuable information, agreement between raters could potentially be high even with significant variability in the actual mean-level ratings. Because of this, this study also examined mean-level comparisons and absolute agreement across ratings. Absolute agreement, as measured by ICCs, tended to be highest across ratings for neuroticism and extraversion. The high agreement on neuroticism might be incongruent with the theory that more internalized, less observable traits would have less agreement between raters (Vazire, 2010). However, this is likely less surprising in the therapist-client relationship in which distress and one's ability to cope with distress is in frequent discussion throughout most sessions.

Because domain-level ICCs only consisted of six items, the full measure ICCs were also examined. Similarly, there were no notable increases across treatment.

An important question regarding agreement was not just how well the clients agreed, but how did this agreement change over time. According to the hierarchical model examined, there were no significant linear trends towards increases or decreases in absolute agreement in this sample, which goes against this study's hypothesis. To investigate any potential domain-level differences in agreement, bootstrapping analyses were also conducted to compare domain ICC scores over time. These results indicated that there were significant decreases on agreement for neuroticism from the first two ratings to ratings three and four.

It is also worth highlighting that agreement between the two raters were not impacted by whether the client viewed themselves in a more negative manner. A caveat to these findings, however, is that the current sample only had nine clients with a *T*-score that would be indicative of an exaggerated negative impression of themselves (i.e., 73 or greater); thus, there was generally a restricted range of scores in this sample. A sample with a more normally distributed range might have seen different relationships to agreement.

Comparison of the ICCs across ratings provided meaningful information regarding client-therapist agreement. Mean-level examinations were able to provide an even richer understanding of the changes in client-therapist agreement over time. It was hypothesized that, on average, clients would report higher levels of traits compared to therapists' average ratings. This was largely supported for openness to experience, such that the largest mean-level discrepancies across ratings were for this domain. Clients, on average, tended to rate themselves as significantly higher on openness to experience compared to the therapists' average ratings. Facet-level analyses suggested that clients repeatedly rated themselves as much more self-aware, broad-minded, and creative compared to therapist ratings, even after attending months of treatment with their therapist. Interestingly, this pattern was recently seen in a meta-analysis by Kim, Di Domenico, and Connelly (2019), such that individuals more generally tended to report themselves as higher on openness to experience compared to informants. This is a particularly noteworthy point, as Bucher and colleagues' (2019) meta-analysis found openness to experience important to a variety of therapeutic outcomes, including working alliance, coping ability, self-rated confidence, and interpersonal improvement. The meta-analysis only examined client-reported personality ratings, however. Based on recent work finding that client- and therapist-report can differentially predict

therapeutic outcomes (Samuel et al., 2018), questions continue to remain regarding who is more accurate in describing the clients' personality.

In other words, if a client sees themselves as more open – even if their therapist does not – do they still receive the same benefits as a client “truly” high in openness to experience? While difficult to truly parse apart the “truth” in these situations, it might be helpful to incorporate a third-party informant, such as a close friend or family member. If the informants' ratings are closer to the clients', it might suggest that the therapist is failing to pick up on meaningful information regarding the clients' personality functioning. On the other hand, if the informants' scores are closer to the therapists' ratings, it might suggest the client has a lack of insight into their current functioning. It might also be worth examining clients' metaperceptions of their personality functioning. Research has shown differences in scores when targets are asked to rate how they see their personality versus how they think others see their personality (Oltmanns & Turkheimer, 2006). In this case, clients might acknowledge that others – including their therapists – see themselves as slightly lower in openness to experience.

Another pattern observed was that of mean-level discrepancies for neuroticism. Although clients and therapists appeared to be in high agreement at the first rating regarding clients' general levels of neuroticism, clients rated themselves as lower on neuroticism compared to their therapists at the three subsequent ratings. Interestingly, in Roberts and colleagues' (2017) meta-analysis, a similar pattern was found for client-reported changes in neuroticism, such that in the studies examined, clients reported significant – and quick – decreases in therapy. The difference in patterns of reported change across the therapist and client here highlight the importance of including a second rater, such as the therapist, as another way to track adaptive changes. That is, while clients might be feeling immediate relief after learning new therapy/coping skills, the therapist might be less willing to see those – perhaps state-level changes – as actual trait-level changes. On the other hand, it could be that therapists are not picking up on the quick change that is truly occurring in clients within the first month or two of treatment.

Relatedly, it is also worth highlighting the difference in overall response styles across therapists and clients. As can be seen when examining differences in standard deviations as well as the spaghetti plots in Figure 2, clients tended to be much more willing to use the extreme endpoints of the FFMRF. The FFMRF was designed to tap into high and low poles of any given personality trait, both of which might be considered maladaptive. For example, someone who

scored high on the facet self-consciousness could be described as timid or embarrassed, whereas someone who scored low could be defined as shameless or glib. It appeared that clients were more willing to describe themselves as very low or very high on specific personality traits, whereas therapists tended to prefer staying around the midpoint. A similar pattern of differing response styles was seen in Samuel and colleagues (2018), such that on the PID-5, therapists were much more hesitant to endorse any level of symptoms related to psychoticism, whereas clients were much more willing to do so. It might be that therapists see these endorsements as more maladaptive than their clients. Taken together, these findings highlight the importance of future research specifically examining whether clients and therapists use personality measures in the same way, such as through measurement invariance. If findings suggest that clients and therapists are *not* using these measures in the same way, there are limitations to assumptions that could be drawn when making comparisons between the two groups.

Client-Therapist Agreement as Predictors of Treatment Outcomes

Across all regression analyses, agreement did not necessarily predict greater engagement or symptoms improvement, especially when using the alpha cut off of .01 as was used in the current study. Thus, the overall findings suggest that the clients' and therapists' agreement with one another – or lack thereof – did not significantly impact the clients' ability to improve throughout sessions. It also did not predict whether or not the client was engaged in treatment by attending sessions more frequently. Rather, it appears that, even if a client and therapist might not conceptualize the client similarly, they can still create a working relationship that keeps the client engaged and fosters improvement. This is particularly important when considering the differences in scores on openness to experience and neuroticism, such that clients tended to rate themselves on more adaptive levels than therapists. Nguyen, Kim, Romain, Tabani, and Chaplin (2020) found that decreases in client-reported neuroticism predicted treatment progress. Based on those findings with that of the current study, it might be that clients' perceived change – whether the therapist sees it as occurring or not – is a protective factor for treatment and can aid in overall improvement. A caveat to Nguyen and colleagues' findings, however, is that both personality ratings and treatment progress were reported only by the client. Future research should look at the relationship between change in traits, particularly decreases in neuroticism, and objective treatment progress variables as well as treatment progress rated by the therapist.

It is important to note that engagement in this study was conceptualized as the client attending many sessions with fewer cancellations, no-shows, or reschedules. However, this is likely only a subset of what “engagement” in therapy truly is. In fact, Tetley, Jinks, Huband, and Howells (2011) described four additional components of active engagement in individual therapy, including fully completing the prescribed treatment within the expected timeframe, completion of home practice between sessions, actively contributing to the treatment process, and developing a working alliance with the therapist. Future research could examine how the agreement between clients and therapists impact these additional components of engagement.

Clinical Implications

The ways in which personality assessment can benefit psychological treatment has been discussed for almost thirty years (e.g., Bagby et al., 2016; Harkness & Lilienfeld, 1997; Mullins-Sweatt & Lengel, 2012; Widiger & Presnall, 2013). Personality ratings *do* provide valuable information regarding treatment outcomes (Bucher et al., 2019), and the information provided can be different based on who is making those ratings (Samuel, Bucher, & Suzuki, 2018). The findings from this study further highlights that, although clients might report quick decreases in levels of neuroticism while receiving treatment, this might be more representative of the clients’ decreases in distress rather than trait neuroticism.

Thus, it is recommended that personality assessments are utilized in clinical settings by both clients, therapist, and perhaps a third informant if possible, to aid in a more holistic understanding of the client. This is particularly important as the field considers and develops ways to specifically target personality traits through treatment interventions, such as through the unified protocol (Allen, McHugh, & Barlow, 2008) and behavioral activation (Magidson, Roberts, Collado-Rodriguez, & Lejuez, 2014). It will also be imperative to continue this line of research as the field moves towards a dimensional conceptualization of personality pathology. Future research should consider the longitudinal assessment and client-therapist agreement for traits on the Alternative Model for Personality Disorders (AMPD) within the *DSM-5*. Although instances of lack of agreement on the FFMRF did not significantly predict any difficulties with client improvement or engagement in this sample, significant disagreement on maladaptive personality traits could potentially have larger negative impacts on these outcomes.

Limitations and Future Directions

The current study was the first to examine repeated assessments of personality traits among clients *and* their therapists. While it provided valuable information, it is not without limitations. A primary limitation was sample size and attrition. The data collected in this study were part of a naturalistic sample, and, over the course of almost nine years, consisted of 128 dyads. Meta-analytic findings have shown dropout rates average around 20%, but they have been found to be as high as 74% (Swift & Greenberg, 2012). Dropout rates tended to be higher in treatments for which there were no predetermined time limits and were conducted in university psychology training clinics. Taken together, the rates of attrition seen in this study are within what would be expected, but they still resulted in limitations regarding conclusions that could be made after more than four ratings. To increase power while dealing with client dropout, attrition, and/or termination, future studies should consider incorporating more frequent assessments, such as weekly rather than monthly. This would also allow for the examination of more nuanced changes occurring throughout treatment. More frequent assessments might be particularly important, given findings that personality changes start to happen as quickly as four weeks into treatment and level off around eight weeks of treatment. Within this study, rating two is on average 6-7 weeks (although most commonly four weeks) and rating three is on average 15 weeks (although most commonly 10). Thus, there might be more nuanced changes occurring between the monthly assessment points in this sample. Relatedly, it is worth noting that the present multilevel analyses examined *linear* personality trajectories and change. Future research might be able to better conceptualize trends in both components via nonlinear means (e.g., Lindstrom & Bates, 1990).

Another limitation to this study was that, because of the naturalistic nature of these data, the rating time points were not unified across all clients. For example, range in days since intake for the first rating went from 0 (i.e., the day of the intake session) to 71 days (almost 10 weeks) after the intake session. The modal number of days after intake for each rating, however, are on par with what would be expected if ratings were completed every four weeks (see Table 2) suggesting that the majority of clients completed these ratings within a more uniformed time. Nonetheless, because of the ranges in days, time was conceptualized as days since the intake session rather than the rating time point in the mixed effect models. Analyses that did rely on looking at each rating time point, such as mean-level comparisons and rank-order agreement, are

likely reflective of the corresponding month of treatment but should be examined with slightly more caution.

An experimental study that included a waitlist control group might better be able to explain the specific mechanisms involved in client-therapist agreement and personality change over time. As stated previously, some experimental studies have already examined interventions that target personality trait change. An extension of the current literature would be incorporating the use of monthly or weekly repeated assessments from both the clients and therapists with a waitlist control group to examine pattern of changes. It would also be worth controlling for state-level distress in such an experiment.

Lastly, the trajectories and slopes of client and therapist ratings in this sample were quite varied and are suggestive of considerable heterogeneity. One source of heterogeneity might be the clients' diagnostic difficulties. Roberts and colleagues (2017) found that change patterns looked different across diagnostic populations, such that clients with anxiety and personality disorders reported greater changes in treatment compared to those presenting with depressive, eating, and substance use disorders, as well as compared to those with comorbid diagnoses, such as the clients commonly seen in the present study (and in the majority of real-life clinical settings). Additionally, past research has shown that how one rates their own level of personality functioning can be influenced by increased levels of personality pathology (Carlson & Oltmanns, 2015). Taken together, it would be imperative to investigate whether these patterns of agreement and trait change look different across diagnostic populations.

Summary

Personality disorders are highly prevalent in clinical settings, and research has suggested that, even with the absence of a personality disorder diagnosis, personality functioning can impact various aspects of treatment. Because of this, the present study sought to examine the longitudinal trajectory of both clients' and therapists' ratings of clients' personality as well as the agreement between the two raters throughout treatment. Results showed significant decreases in neuroticism that were only unique to the clients' ratings, suggesting that clients reported increases in emotional stability during the first months of treatment that therapists did not report observing. However, it is worth noting that clients' self-reported levels of distress changed at similar patterns as clients' self-reported levels of neuroticism. Given this as well as the association between distress and

neuroticism trait scores, it is likely that clients' ratings on neuroticism were largely influenced by their state-level distress. On the other hand, therapists continually rated clients based on their perception of the clients' trait-level functioning of neuroticism. Although absolute agreement – or lack thereof – did not impact treatment outcomes related to engagement and symptom improvement, the findings highlight the importance of using a second rater, such as a therapist, when tracking treatment goals and symptom change. Overall, the findings further elucidate personality trait change across intervention from multiple raters and assessments while incorporating state-level distress's impact on trait change. Future research should continue to examine trait change and client-therapist agreement in both experimental and naturalistic settings and with both adaptive and maladaptive personality traits.

REFERENCES

- Achenbach, T. M., Krukowski, R. A., Dumenci, L., & Ivanova, M. Y. (2005). Assessment of adult psychopathology: Meta-analyses and implications of cross-informant correlations. *Psychological Bulletin*, 131(3), 361-382. <https://doi.org/10.1037/0033-2909.131.3.361>
- Allan, J., Leeson, P., Fruyt, F. De, & Martin, S. (2018). Application of a 10 week coaching program designed to facilitate volitional personality change: Overall effects on personality and the impact of targeting. *International Journal of Evidence Based Coaching and Mentoring*, 16(1), 80-94. <https://doi.org/10.24384/000470>
- Allemand, M., & Flückiger, C. (2017). Changing personality traits: Some considerations from psychotherapy process-outcome research for intervention efforts on intentional personality change. *Journal of Psychotherapy Integration*, 27(4), 476-494. <https://doi.org/10.1037/int0000094>
- Allen, L. B., McHugh, R. K., & Barlow, D. H. (2008). Emotional disorders: A unified protocol. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders: A step-by-step treatment manual* (pp. 216-249). New York, NY: The Guilford Press.
- Allen, M. S., Vella, S. A., & Laborde, S. (2015). Health-related behaviour and personality trait development in adulthood. *Journal of Research in Personality*, 59, 104-110. <https://doi.org/10.1016/j.jrp.2015.10.005>
- Bagby, R. M., Gralnick, T. M., Al-Dajani, N., & Uliaszek, A. A. (2016). The role of the Five-Factor Model in personality assessment and treatment planning. *Clinical Psychology: Science and Practice*, 23(4), 365-381.
- Bateman, A., & Fonagy, P. (2008). 8-Year follow-up of patients treated for borderline personality disorder: Mentalization-based treatment versus treatment as usual. *American Journal of Psychiatry*, 165(5), 631-638. <https://doi.org/10.1176/appi.ajp.2007.07040636>
- Bernstein, D. P., Iscan, C., & Maser, J. (2007). Opinions of personality disorder experts regarding the DSM-IV personality disorders classification system. *Journal of Personality Disorders*, 21(5), 536-551. <https://doi.org/10.1521/pedi.2007.21.5.536>

- Bleidorn, W., Hopwood, C. J., Ackerman, R. A., Witt, E. A., Kandler, C., Riemann, R., Samuel, D. B., & Donnellan, M. B. (2019). The healthy personality from a basic trait perspective. *Journal of Personality and Social Psychology*, 118(6), 1207-1225. <https://doi.org/10.1037/pspp0000231>
- Bucher, M. A., Suzuki, T., & Samuel, D. B. (2019). A meta-analytic review of personality traits and their associations with mental health treatment outcomes. *Clinical Psychology Review*, 70, 51-63. <https://doi.org/10.1016/j.cpr.2019.04.002>
- Carlson, E. N., & Oltmanns, T. F. (2015). The role of metaperception in personality disorders: Do people with personality problems know how others experience their personality? *Journal of Personality Disorders*, 29(4), 449-467.
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136(6), 1092-1122. <https://doi.org/10.1037/a0021212>
- Costa, P. T., Bagby, R. M., Herbst, J. H., & McCrae, R. R. (2005). Personality self-reports are concurrently reliable and valid during acute depressive episodes. *Journal of Affective Disorders*, 89(1-3), 45-55. <https://doi.org/10.1016/j.jad.2005.06.010>
- Crawford, M., Rutter, D., Price, K., Weaver, T., Jossion, M., Tyrer, P., Gibson, S., Gillespie, S., Faulkner, A., Ryrie, I., Dhillon, K., Bateman, A., Fonagy, P., Taylor, B., Moran, P., Beckett, J., Blackwell, H., Burbridge, C., Coldham, T., . . . Taylor, E. (2007). *Learning the lessons: A multi-method evaluation of dedicated community-based services for people with personality disorder*. Report for the National Co-ordinating Centre for NHS Service Delivery and Organisation R&D (NCCSDO). London, England: Queen's Printer and Controller of HMSO 2007.
- Eisman, E. J., & Nordal, K. (2017). The implications for health care: Commentary on the special section on teaching, training, and supervision in personality and psychological assessment. *Journal of Personality Assessment*, 99(2), 186-188. <https://doi.org/10.1080/00223891.2016.1226176>
- Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., Audin, K., Connell, J., & McGrath, G. (2000). CORE: Clinical Outcomes in Routine Evaluation. *Journal of Mental Health*, 9(3), 247-255. <https://doi.org/10.1080/jmh.9.3.247.255>

- Glover, N. G., Crego, C., & Widiger, T. A. (2012). The clinical utility of the Five Factor Model of personality disorder. *Personality Disorders: Theory, Research, and Treatment*, 3(2), 176-184. <https://doi.org/10.1037/a0024030>
- Griens, A. M. G. F., Jonker, K., Spinhoven, P., & Blom, M. B. J. (2002). The influence of depressive state features on trait measurement. *Journal of Affective Disorders*, 70(1), 95-99. [https://doi.org/10.1016/S0165-0327\(00\)00371-2](https://doi.org/10.1016/S0165-0327(00)00371-2)
- Harkness, A. R., & Lilienfeld, S. O. (1997). Individual differences science for treatment planning: Personality traits. *Psychological Assessment*, 9(4), 349-360.
- Hudson, N. W., & Chris Fraley, R. (2015). Volitional personality trait change: Can people choose to change their personality traits? *Journal of Personality and Social Psychology*, 109(3), 490-507. <https://doi.org/10.1037/pspp0000021>
- Human, L. J., Biesanz, J. C., Miller, G. E., Chen, E., Lachman, M. E., & Seeman, T. E. (2013). Is change bad? Personality change is associated with poorer psychological health and greater metabolic syndrome in midlife. *Journal of Personality*, 81(3), 249-260. <https://doi.org/10.1111/jopy.12002>
- Kim, H., Di Domenico, S. I., & Connelly, B. S. (2019). Self-other agreement in personality reports: A meta-analytic comparison of self- and informant-report means. *Psychological Science*, 30(1), 129-138. <https://doi.org/10.1177/0956797618810000>
- Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine*, 42, 1879-1890.
- Lambert, M. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, 17(1), 1-14. <https://doi.org/10.1080/10503300601032506>
- Lindstrom, M. J., & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46(3), 673-687.
- Linehan, M. M., Cochran, B. N., Mar, C. M., Levensky, E. R., & Comtois, K. A. (2000). Therapeutic burnout among borderline personality disordered clients and their therapists: Development and evaluation of two adaptations of the Maslach Burnout Inventory. *Cognitive and Behavioral Practice*, 7(3), 329-337. [https://doi.org/10.1016/S1077-7229\(00\)80091-7](https://doi.org/10.1016/S1077-7229(00)80091-7)

- Loranger, A. W., Lenzenweger, M. F., Gartner, A. F., Susman, V. L., Herzig, J., Zammit, G. K., Gartner, J. D., Abrams, R. C., & Young, R. C. (1991). Trait-state artifacts and the diagnosis of personality disorders. *Archives of General Psychiatry*, 48(8), 720-728. <https://doi.org/10.1001/archpsyc.1991.01810320044007>
- Magidson, J. F., Roberts, B. W., Collado-Rodriguez, A., & Lejuez, C. W. (2014). Theory-driven intervention for changing personality: Expectancy value theory, behavioral activation, and conscientiousness. *Developmental Psychology*, 50(5), 1442-1450. <https://doi.org/10.1037/a0030583>
- McMurran, M., Huband, N., & Overton, E. (2010). Non-completion of personality disorder treatments: A systematic review of correlates, consequences, and interventions. *Clinical Psychology Review*, 30(3), 277-287. <https://doi.org/10.1016/j.cpr.2009.12.002>
- Morey, L. C. (2007). *The personality assessment inventory professional manual* (2nd ed.). Lutz, FL: Psychological Assessment Resources.
- Mullen, L. S., Blanco, C., Vaughan, S. C., Vaughan, R., & Roose, S. P. (1999). Defense mechanisms and personality in depression. *Depression and Anxiety*, 10(4), 168-174. [https://doi.org/10.1002/\(SICI\)1520-6394\(1999\)10:4<168::AID-DA5>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1520-6394(1999)10:4<168::AID-DA5>3.0.CO;2-5)
- Mullins-Sweatt, S. N., Jamerson, J. E., Samuel, D. B., Olson, D. R., & Widiger, T. A. (2006). Psychometric properties of an abbreviated instrument of the five-factor model. *Assessment*, 13(2), 119-137. <https://doi.org/10.1177/1073191106286748>
- Mullins-Sweatt, S. N., & Lengel, G. J. (2012). Clinical utility of the Five-Factor Model of Personality Disorder. *Journal of Personality*, 80(6), 1615-1639. <https://doi.org/10.1111/j.1467-6494.2012.00774.x>
- Newton-Howes, G., Tyrer, P., Anagnostakis, K., Cooper, S., Bowden-Jones, O., & Weaver, T. (2010). The prevalence of personality disorder, its comorbidity with mental state disorders, and its clinical significance in community mental health teams. *Social Psychiatry and Psychiatric Epidemiology*, 45, 453-460. <https://doi.org/10.1007/s00127-009-0084-7>
- Nguyen, P. L. L., Kim, H. L., Romain, A. M. N., Tabani, S., & Chaplin, W. F. (2020). Personality change and personality as predictor of change in psychotherapy: A longitudinal study in a community mental health clinic. *Journal of Research in Personality*, 87, 103980. <https://doi.org/10.1016/j.jrp.2020.103980>

- Oltmanns, J. R., & Oltmanns, T. F. (2021). Self–other agreement on ratings of personality disorder symptoms and traits: Three meta-analyses. In T. D. Letzring & J. S. Spain (Eds.), *The handbook of accurate personality judgment: Theory and empirical findings*. Oxford University Press.
- Oltmanns, T. F., & Turkheimer, E. (2006). Perceptions of self and others regarding pathological personality traits. In R. F. Krueger & J. L. Tackett (Eds.), *Personality and Psychopathology* (pp. 71-111). New York, NY: Guilford Press.
- Pinheiro, J., Bates, D. M., DebRoy, S., & Sarkar, D. (2021). nlme: Linear and nonlinear mixed effects models.
- Reynolds, S. K., & Clark, L. A. (2001). Predicting dimensions of personality disorder from domains and facets of the five-factor model. *Journal of Personality*, 69(2), 199-222. <https://doi.org/10.1111/1467-6494.00142>
- Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin*, 143(2), 117-141.
- Roberts, B. W., & Mroczek, D. (2008). Personality trait change in adulthood. *Current Directions in Psychological Science*, 17(1), 31-35. <https://doi.org/10.1111/j.1467-8721.2008.00543.x>
- Rosenberg, E. L. (1998). Levels of analysis and the organization of affect. *Review of General Psychology*, 2(3), 247-270. <https://doi.org/10.1037/1089-2680.2.3.247>
- Samuel, D. B. (2015). A review of the agreement between clinicians' personality disorder diagnoses and those from other methods and sources. *Clinical Psychology: Science and Practice*, 22(1), 1-19. <https://doi.org/10.1111/cpsp.12088>
- Samuel, D. B., Bucher, M. A., & Suzuki, T. (2018). A preliminary probe of personality predicting psychotherapy outcomes: Perspectives from therapists and their clients. *Psychopathology*, 51, 122-129. <https://doi.org/10.1159/000487362>
- Samuel, D. B., Suzuki, T., Bucher, M. A., & Griffin, S. A. (2018). The agreement between clients' and their therapists' ratings of personality disorder traits. *Journal of Consulting and Clinical Psychology*, 86(6), 546-555. <https://doi.org/10.1037/ccp0000304>
- Samuel, D. B., & Widiger, T. A. (2008). A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: A facet level analysis. *Clinical Psychology Review*, 28, 1326-1342.

- Samuels, J. (2011). Personality disorders: Epidemiology and public health issues. *International Review of Psychiatry*, 23(3), 223-233. <https://doi.org/10.3109/09540261.2011.588200>
- Sauer-Zavala, S., Fournier, J. C., Steele, S. J., & Woods, B. K. (2020). Does the unified protocol really change neuroticism? Results from a randomized trial. *Psychological Medicine*, 21, 1-10.
- Soskin, D. P., Carl, J. R., Alpert, J., & Fava, M. (2012). Antidepressant Effects on Emotional Temperament: Toward a Biobehavioral Research Paradigm for Major Depressive Disorder. *CNS Neuroscience and Therapeutics*, 18(6), 441-451. <https://doi.org/10.1111/j.1755-5949.2012.00318.x>
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*, 100(2), 330-348. <https://doi.org/10.1037/a0021717>
- Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 80(4), 547-559. <https://doi.org/10.1037/a0028226>
- Tang, T. Z., DeRubeis, R. J., Hollon, S. D., Amsterdam, J., Shelton, R., & Schalet, B. (2009). Personality change during depression treatment: A placebo-controlled trial. *Archives of General Psychiatry*, 66(12), 1322-1330. <https://doi.org/10.1001/archgenpsychiatry.2009.166>
- Tetley, A., Jinks, M., Huband, N., & Howells, K. (2011). A systematic review of measures of therapeutic engagement in psychosocial and psychological treatment. *Journal of Clinical Psychology*, 67(9), 927-941. <https://doi.org/10.1002/jclp.20811>
- Turiano, N. A., Pitzer, L., Armour, C., Karlamangla, A., Ryff, C. D., & Mroczek, D. K. (2012). Personality trait level and change as predictors of health outcomes: Findings from a national study of Americans (MIDUS). *Journals of Gerontology - Series B Psychological Sciences and Social Sciences, Series B*, 67B(1), 4-12. <https://doi.org/10.1093/geronb/gbr072>
- Vazire, S. (2010). Who knows what about a person? The Self-Other Knowledge Asymmetry (SOKA) Model. *Journal of Personality and Social Psychology*, 98(2), 281-300. <https://doi.org/10.1037/a0017908>

- Widiger, T.A., & Costa, P. T. (1994). Personality and personality disorders. *Journal of Abnormal Psychology, 103*, 78-91.
- Widiger, Thomas A, & Presnall, J. R. (2013). Clinical application of the Five-Factor Model. *Journal of Personality, 81*(6), 515-527.
- Wrzus, C., & Roberts, B. W. (2016). Processes of personality development in adulthood: The TESSERA Framework. *Personality and Social Psychology Review, 21*(3), 253-277.
<https://doi.org/10.1177/1088868316652279>