PERSONALITY ATTRIBUTES IN CLINICAL PRESENTATION, MEASUREMENT, AND TREATMENT

by

AMBER GAYLE THALMAYER

A DISSERTATION

Presented to the Department of Psychology
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2013

DISSERTATION APPROVAL PAGE

Student: Amber Gayle Thalmayer

Title: Personality Attributes in Clinical Presentation, Measurement, and Treatment

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Psychology by:

Gerard Saucier	Chairperson
Sanjay Srivastava	Core Member
Don Tucker	Core Member
Joseph Stevens	Institutional Representative

and

Kimberly Andrews Espy	Vice President for Research and Innovation; Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded September 2013

DISSERTATION ABSTRACT

Amber Gayle Thalmayer

Doctor of Philosophy

Department of Psychology

September 2013

Title: Personality Attributes in Clinical Presentation, Measurement, and Treatment

Psychotherapy is sought for diverse problems, and trust in its efficacy has led to increasing parity in insurance coverage for psychological services. But about half of those who begin therapy drop out prematurely, and only about half of those who complete therapy experience significant improvements. Here issues of efficacy measurement and the potential role of personality differences to better guide and assess treatment are explored. It is hoped this knowledge could lead to increased success rates.

In terms of outcome measurement, the use and psychometric properties of the Outcome Questionnaire-45 (OQ-45), a popular, brief measure of psychological functioning and change, are assessed. Factor structure is tested in clinical (Study 1) and student (Study 2) samples. Alternative scoring models test whether dimensions of personality drive responses. Using bi-level models, including a total score factor, fit of the intended structure and three- and four-factor personality models was similar. A seven-factor problems model provided the best fit. About half the variance in OQ-45 scores was accounted for by a Big Six personality inventory. The best items for assessing personality attributes in OQ-45 data are noted.

Secondly, the relation between personality attributes and clinical presentation and outcomes is explored. Self-report scores on personality attributes predict virtually every

life outcome and are highly associated with clinical presentation. Such attributes surely also play a role in treatment, but this association has not been widely studied. In Study 3 the relation of personality attributes to presentation, usage, and outcome is tested in a Couples and Family Therapy clinic sample (N = 222). Neuroticism was strongly associated with initial OQ-45 score, as were other scales (except Openness) to a lesser degree. Being older, more educated, and married predicted attending more sessions, and being older, female, and higher income predicted termination success. Personality scales did not play a role in either outcome. Honesty/Propriety and Neuroticism were, however, associated with steeper decrease in OQ-45 scores over time.

CURRICULUM VITAE

NAME OF AUTHOR: Amber Gayle Thalmayer

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Southern Oregon University, Ashland
Goddard College, Plainfield, Vermont

DEGREES AWARDED:

Doctor of Philosophy, Psychology, 2013, University of Oregon

Master of Science, Psychology, 2008, University of Oregon

Bachelor of Science, Psychology, minor in German Language and Literature, Southern Oregon University

AREAS OF SPECIAL INTEREST:

Personality structure and measurement
Cross-cultural issues in personality
Mental health treatment

PROFESSIONAL EXPERIENCE:

Course instructor, Department of Psychology, University of Oregon, Eugene, 2008 - 2013

Teaching assistant and lab instructor, Department of Psychology, University of Oregon, Eugene, 2006 - 2013

Research Assistant, Department of Psychology, 2006 - 2012

Neuropsychological Testing Assistant and Interviewer, Dr. Steve Pethick, Eugene, Oregon, 2010 – 2013

Group Facilitator and diagnostic interviewer, Oregon Research Institute, Eugene, Oregon, 2011, 2013

Data Analyst, Decision Research Institute, Eugene, Oregon, 2009

GRANTS, AWARDS, AND HONORS:

Graduate Teaching Fellowship, UO Department of Psychology, 2006 to present

Betty Foster McCue Fellowship for Human Performance and Development, UO Graduate School 2012

Carolyn Stokes Memorial Scholarship, UO College of Arts and Sciences, 2012 & 2010

Gregores Award, UO Department of Psychology, 2012

Clarence and Lucille Dunbar Scholarship, UO College of Arts and Sciences, 2011

Norm Sundberg Award, UO Department of Psychology, 2011

American Association of University Women, Outstanding Graduating Woman in Psychology 2006

Summa cum laude, Southern Oregon University, 2006

PUBLICATIONS:

Thalmayer, A.G. Personality attributes and the Outcome Questionnaire-45. Manuscript submitted for publication.

Saucier, G., Thalmayer, A.G., & Bel-Bahar, T. Personality descriptors ubiquitous across 12 languages. Manuscript submitted for publication.

Saucier, G., Thalmayer, A.G., Payne, D. L., Carlson, R., Sanogo, L., Ole-Kotikash, L., Church, A.T., Katigbak, M. S., Szarota, P., Szirmák, Z., & Zhou, X. (in press). A basic bivariate structure of personality attributes evident across nine languages. *Journal of Personality*.

Saucier G. & Thalmayer, A. G. (2013). Cultural, contextual, and measurement challenges for the paradigm of personality science. In Christiansen, N. & Tett, R. (Eds) *Handbook of personality at work.* New York: Taylor & Francis Psychology Press.

Thalmayer, A.G., Saucier, G., & Eigenhuis, A. (2011). The comparative validity of brief to medium-length Big Five and Big Six personality questionnaires. *Psychological Assessment, 23,* 995-1009.

ACKNOWLEDGMENTS

Thank you to my advisor Gerard Saucier for many enriching opportunities, and for helping me learn to think scientifically and with increased depth and breadth. Thank you to committee members Joseph Stevens, Sanjay Srivastava and Don Tucker for their support and assistance with my dissertation and with many other projects. My dissertation studies would not have been possible without the support of Couples and Family Therapy clinic directors Tiffany Brown and Jeff Todhal, who allowed me to collect data and work with existing data in their archives, and without ongoing assistance from clinic staff members Lalla Pudewell, Lindsay Brown, and Caitlin Peoples. This work also depended on the hard work of several undergraduate research assistants, in particular Nathan Baune, as well as Grant Rusin, Brenna Bell, Kristen Ackerman, and Thomas Bradley. Thank you to fellow doctoral students Annemarie Eigenhuis and Zhuo Job Chen for practical advice on aspects of this project, and to them and many others, in particular Ida Moadab, Allison Waters, Philippe Bou Malham, and Kathryn Iurino for camaraderie. I am grateful to many family members and close friends, especially Merilyn Wakefield, Stacy Wakefield Forte, and Elaine Deckelman, for their ongoing enthusiasm, perspective, and support. For sustaining encouragement and support virtually every day of this seven year process, I most of all thank my husband, Erich Mead Thalmayer with whom this accomplishment is shared.

Dedicated, in memoriam, to my father, John Arthur Wakefield -- a man who revered intellectual achievements and who had more than the usual amount of personality. I think he would have been the proudest of all to have a Ph.D. in the family.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

OVERVIEW

Psychotherapy is sought and recommended for a wide range of problems, and trust in its efficacy has led to increasing parity in insurance coverage for psychological services. We know that evidence-supported therapies work for a small majority of patients with common disorders (depression, anxiety, eating disorders). Many aspects of clients, therapist behaviors, and treatment mechanisms are being studied in order to improve the rate of success. In my dissertation research I will examine issues of both efficacy and individual differences. I will first explore questions of optimal outcome measurement -- how can efficacy best be quantified and tracked? Secondly, I will address how measurable aspects of individual differences contribute to therapy outcomes. Can robust personality and temperament constructs be used to better guide and assess treatment? Ultimately I hope to find ways to make therapy successful for a more diverse array of people by exploring the role of relatively stable behavioral tendencies in therapy usage and outcomes.

Outcome measurement here focuses on the effects of general counseling. In some cases when an individual presents for therapy, a diagnosis of a specific mental disorder is made. Established protocols address some specific diagnoses, and there are instruments (such as the Beck Depression Inventory) designed to capture change in specific symptoms during the treatment of a specific disorder. This model of diagnosis and protocol driven treatment is homologous to the medical model and lends itself well to clinical trials, which typically include only people who meet criteria for a specific Diagnostic Statistical Manual (DSM; APA, 2000) disorder. However, the majority of therapy that is administered is not protocol-driven. This may be due to a variety of reasons. First, even the best-validated therapies for specific disorders generally only help about half of clients; thus, the verdict is still out on ideal treatment. Second, many who present for therapy do not meet criteria for a diagnosis, or may meet criteria for several. Third, some argue that the evidence for the efficacy of psychotherapy does not support the individualistic and mechanistic paradigm of patients as “carriers” of disorders and therapists as “more or less skilled at administering the optimal procedures for each diagnosed disorder” but instead that therapeutic efficacy in based in the patient’s

experience of a "remoralizing, resource enhancing, and motivating relationship with a therapist…" (Orlinsky, 2010, p. xxi). Finally, others may see different types of therapy not as competitors, but as appropriately diverse, given the vast range of persons and types of problems that mental health professionals attempt to treat. Regardless of whether for practical reasons (a delay in dissemination) or because protocol-driven therapies have specific domains of efficacy that do not encompass the majority of those who seek mental health counseling services, most clients receive more general counseling services that are not based on a the treatment of a specific disorder with a specific protocol (Shafran, et al., 2009).

The diversity of clients and presenting issues in such settings, and the practical reality of limited resources (the need to dive quickly into treatment rather than devote many hours to comprehensive assessment) puts many constraints on measurement. However, some form of quantitative tracking is necessary in order to provide clinics, therapists, and insurers with basic information about the extent of problems in functioning, and the extent and rate of improvement. How to maximize the usefulness of measurement (the best instrument for the job, and the best use of results) in real world settings is of great importance.

In terms of therapy fit, a body of clinical research has focused on the fit between specific diagnoses and specific protocols. However, as noted above, even the best-validated procedures appear to work for only about half of the clients who receive them. I aim to explore how easily measurable client variables, including personality attributes, relate to therapy outcomes. Ultimately this preliminary research might aid in an effort to help direct patients to the type of therapy that will be most effective for them.

The first part of the literature review examines the clinical use and psychometric properties of the Outcome Questionnaire-45 (OQ-45). This popular, brief instrument is designed to cheaply and efficiently measure level of functioning and track change in the wide variety of clients typically seen in a community clinic. The design, convergent and discriminant validity of this questionnaire, and its factor structure are reviewed and tested in clinical (Study 1) and college student (Study 2) samples. Alternative scoring models and interpretation of the OQ-45 are compared. Moreover, correlations between OQ-45

items and scales and personality attribute dimensions are tested in the college student sample.

The second part of the literature review assesses the relation between personality attributes and clinical presentation and outcomes. Literature on the relation of Big Five and Big Six personality dimensions to clinical presentation and treatment is reviewed, leading to the conclusion that normal range personality attributes are consistently and meaningfully linked with the former, and logically highly likely to be meaningfully linked with the latter. Study 3 involves data collected in the University of Oregon's Couples and Family Therapy clinic, in which the relations of personality attributes to clinical presentation (initial OQ-45 scores), to three therapy usage and outcome variables (number of sessions attended, termination outcome, and change in OQ-45 scores) are tested.

CHAPTER II

LITERATURE REVIEW

Clinical Use of the Outcome Questionnaire-45: A Psychometric Exploration

History and Design

The Outcome Questionnaire-45 (OQ-45) was developed by Lambert, Morton, and colleagues (2004) not as a diagnostic tool, but as a measure of common symptoms across a range of mental disorders and syndromes, including stress-related illness, to be used as a baseline screening instrument and as a means of measuring change in symptoms over time. Administration of this a questionnaire takes only a few minutes of client time, and allows for a relatively objective quantitative assessment of the effectiveness of services, an assessment that is importantly separate from (but ideally informs) the clinical judgment that occurs within the context of the therapeutic relationship (Lambert, Gregersen, & Burlingame, 2004).

Items for the OQ-45 were rationally selected to assess the most common symptoms across patients and disorders, and symptoms that might most directly affect quality of life in the general population. Reviews of previous literature and an effort to achieve concordance with *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., revised; *DSM-III-R*) diagnoses identified three broad content areas: (1) Symptom Distress (intrapsychic problems including anxiety disorders, mood disorders, and substance-related disorders, e.g., “I feel blue”); (2) interpersonal problems (e.g., “I feel lonely”); and (3) social role functioning (e.g., “I feel stressed at work/school). Items were written with the intention that they would fit into one of these three domains (see Appendix Table for the items, sorted by intended subscale).

The administration manual (Lambert, Morton, et al., 2004) explains that the Symptom Distress category, at 25 items, is the largest because affective disorders are the most commonly diagnosed. It also explains that because recent literature indicates that the symptoms of anxiety and depression cannot easily be distinguished, no attempt was made to separate them in the OQ-45. Indeed while large studies of the occurrence and comorbidity of mental disorders continue to confirm that anxiety and depression often co-occur and can be conceptualized as both belonging to a broader domain of internalizing disorders (e.g. Krueger & Markon, 2006), such studies also generally report bi-furcation

of the internalizing domain into subcategories of misery and fear (Krueger & Markon, 2006). The mixing of these disorder tendencies into one category of the OQ-45 thus indicates some potential for a failure of unidimensionality. More problematically, because "next to these disorders substance abuse was the most common diagnosis," items regarding substance abuse were included on the same dimension (Lambert et al., 2004). However, the same studies (e.g. Krueger & Markon, 2006) that make the case for a broad class of internalizing tendencies, also make a case for a separate domain of externalizing disorders, including substance abuse.

The second subscale, Interpersonal Relations, was designed to capture satisfaction and problems (friction, conflict, isolation, inadequacy, withdrawal) in relationships, due to the well-established association between relationships and well-being, and the frequency of interpersonal problems as topics in therapy. Items were derived from literature on marriage and family therapy and from research on interpersonal problems reported by clients (Lambert, Gregersen, et al., 2004).

The third subscale seeks to measure Social Role performance ("dissatisfaction, conflict, distress, and inadequacy in tasks related to …employment" 2004, p. 2), based on the rationale that psychological problems and work role performance can affect each other bi-directionally (Lambert, Gregersen, et al., 2004).

Convergent and Discriminant Validity

People with more severe mental health presentations consistently receive higher self-report scores on the OQ-45. The correlation of the total score, the Symptom Distress score, and to a lesser extent the other two subscale scores, with clinical severity has been well established (Umphress, Lambert, Smart, Barlow, & Clouse, 1997; de Beurs et al.,2005) and has been found for the OQ-45 in translation to Spanish (von Bergen & de la Parra, 2002), German (Haug et al., 2004), Dutch (De Jong et al., 2007), and Chinese (Qin & Hu, 2008; Li & Luo, 2009).

It also appears that the OQ-45 is sensitive to change, and that total score reliably diminishes as clients' symptom severity decreases in both outpatient and inpatient samples (Doerfler, Addis, & Moran, 2002; von Bergen & de la Parra, 2002; Haug et al., 2004; de Beurs et al 2005; Talley & Clack, 2006).

There is a lack of evidence, however, for the discriminant validity of the three subscales. Umphress and colleagues (1997) provided an early assessment of construct validity for the OQ-45 using community, university counseling center, community clinic, and inpatient samples and found the three OQ-45 subscales to be highly correlated. The Symptom Distress (SD) subscale correlated so highly with total score (.96-.98) as to appear to be synonymous with the higher order construct. And while the Global Severity Index of the Symptom Checklist-90-Revised (SCL-90) correlated highly with SD (.82-.92) and with total score (.78-.88), the two other OQ-45 subscales failed to show the hypothesized pattern of correlation with measures chosen to test their divergent validity (the Inventory of Interpersonal Problems and the Social Adjustment Scale). Likewise Hess, Rohlfing, Hardy, Glidden-Tracey and Tracey (2010) found no evidence that the Interpersonal Relation scale of the OQ-45 had more relation to the Inventory of Interpersonal Problems than any other OQ-45 scale.

Doerfler, Addis, and Moran (2002) compared the OQ-45 to the Behavior and Symptom Identification Scale (BASIS-32), a 32-item, self-report questionnaire designed to assess symptoms and social functioning in psychiatric inpatients in a sample of hospitalized inpatients. Overall BASIS subscales and total score correlated more highly with Symptom Distress than with the OQ total or the other scales. This was true even for BASIS "Relation to Self and Others" and "Daily Living and Role Functioning", scales which logically should have had higher correlations with Interpersonal Relations and Social Role, respectively. The authors concluded that OQ-45 subscales fail to capture distinct aspects of patient functioning. However, this was also true to some extent for BASIS subscales. While this could be due to high comorbidity, rather than shortcomings of specific scales, the evidence suggests that for both measures the total score rather than subscale scores should be used.

There is some initial evidence that OQ-45 scores might differentiate between clinical diagnoses (Doerfler, Addis, & Moran, 2002), although this is not a claim put forward by the OQ authors. In a study of 550 hospitalized inpatients, scores at intake on Symptom Distress were statistically significantly higher for patients who were diagnosed with major depression and bi-polar disorder than for those diagnosed with schizophrenia, although the differences were small. (Higher scores for those with affective disorders

compared to those with schizophrenia were also observed for the Social Role scale, although this difference was not hypothesized. Differences between groups were not observed for the Interpersonal Relations scale, and were not assessed for total score.)

Because the instrument was not designed to diagnose specific disorders, however, the authors of the OQ-45 do not appear to have sought discriminant ability. Thus, in general the OQ total and subscales tend to converge with everything related to psychopathology, but they do not make distinctions. This may not be required of a short measure that reliably and efficiently captures level of functioning across diverse patients in diverse settings.

However, if the subscales do not aid in interpretation, the length of the measure could be reexamined. Length requires client effort, and is sure to attenuate the quantity and quality of completions, especially from individuals in crises or those whose difficulties in functioning make them late to appointments. If the OQ-45 could maintain its power to rate overall level of functioning and to track relevant change, a shorter version would have obvious advantages.

Do Personality Differences Drive OQ-45 Scores?

It is also worth considering whether the OQ-45 may simply function as a general measure of emotional instability, a construct commonly termed Neuroticism or (if referring to its opposite pole) Emotional Stability in personality attribute models like the Big Five. This construct encompasses the tendency to experience more or less activation of internalizing negative emotions (Nettle, 2007), and appears to depend on individual differences in the reactivity of the amygdala and limbic system (Whittle et al, 2006). Robust individual differences in this level of reactivity have a strong genetic influence (Bouchard, 2004) and are visible in infancy (Rothbart, 2007). Neuroticism/Negative Affectivity is thus a relatively stable lifelong trait -- some people will always tend to be a little higher or lower in their reactivity. Personality psychologists understand most psychological disorders, in particular affective disorders, to occur more commonly in persons high on this trait (Clark, 2005; Funder, 2010; Nettle, 2007). Any measure that taps into a general tendency to feel more negative emotions (and to experience associated somatic symptoms) can thus be expected to correlate with clinical status and with level of functioning. To the extent that the OQ-45 is solely functioning as a measure of this broad

trait, it is doing so inefficiently, as the trait might be effectively captured with as few as two items (Rammstedt & John, 2009), and scales with sixteen items may perform indistinguishably from scales with eight (Thalmayer, Saucier, & Eigenhuis, 2011).

The goal of the OQ-45 is not to capture stable trait differences, but more dynamic state differences. However, the distinction between trait and state is blurry, and the two lie on a continuum. In this domain trait scales have been shown to capture state variation (Neuroticism trait scores goes down after treatment for depression; De Fruyt et al., 2006) and state measures such as the SCL, the OQ, BASIS, and the Beck Depression and Anxiety inventories certainly capture underlying individual differences in the stable trait as well as state variation.

Some researchers, noticing this overlap, have called for a more widespread use of well-validated personality attribute inventories in clinical settings (Widiger & Trull, 2007), which can aid in diagnosis and treatment of diverse disorders (Clark, 2009). As yet, no studies have assessed the relation of the OQ-45 to Big Five or Big Six attributes, or tested whether the OQ-45 provides incremental validity beyond what could be assessed by these basic, robust domains of individual psychological differences.

Structure

Since the wide implementation and translation of the OQ-45, studies testing its structure using confirmatory factor analysis (CFA) and related methods have not provided support for the proposed structure. Mueller, Lambert, and Burlingame (1998) compared the intended model to a total score and two-factor model (collapsing second two subscales), and concluded that the three-factor model was slightly better, but no models met standard criteria for acceptable fit. Lack of fit for the intended model has also been reported in Spanish (Jurado, 2007), Dutch (De Jong et al., 2007) and Swedish (Wennberg et al., 2010) versions (cf. Chinese; Li & Luo, 2009).

Adequate fit for the intended model was observed in an Italian version, however, when items were allowed to load on their subscale and on a total score scale (although most variance in IR items was accounted for by total score, and the SR scale had low reliability; Coco et al., 2008). This bi-level solution appears to better match the usage of the OQ-45. Bludworth, Tracey and Glidden-Tracey (2010) found support for this bi-level model in a large American sample of university counseling clients. They observed that

items generally loaded highly and evenly on the total score, with lower loadings on subscales.

The authors of confirmatory studies have used a variety of criteria, and have proposed differing solutions. For example, De Jong and colleagues (2007) found two extra factors (one consisting of social role items and another that reflected anxiety and somatic symptoms) by doing a PCA over the residual matrix after fitting the three-factor model. At the item level, some suggest dropping substance abuse items because they don't fit the structure and are highly skewed (Coco et al., 2008; Kim, Beretvas, & Sherry, 2010), though removal might reduce clinical utility. Item 14 'I work/study too much' fails to correlate with other items, and scores are sometimes higher in non-clinical samples (De Jong et al., 2007). But no studies to date have tested alternative models beyond the OQ-45 subscales (e.g., collapsing SR and IR) -- other latent factors that may drive responses have not been explored.

Some Goals for Research

There are several psychometric problems with the OQ-45. Even an initial consideration of the content of the subscales indicates that responses on the measure will be unlikely to fit the proposed structure of the questionnaire, in particular due to the presence of items measuring depression, anxiety and substance abuse all on the Symptom Distress subscale. Studies of the convergent and divergent validity of the OQ fail to find evidence for the construct validity of the three subscales: the scales correlate highly with each other; Symptom Distress appears not to be meaningfully distinct from the total score; and Social Role and Interpersonal Relation subscales show no special pattern of relation to other instruments with which they should logically converge (Umphress et al., 1997; Hess et al., 2010). Thus, unsurprisingly, confirmatory factor analyses using a variety of samples in a variety of languages have failed to support the intended structure of the instrument (Mueller et al., 1998; Jurado, 2007; De Jong et al., 2007; Wennberg et al., 2010). The results of these studies suggest that the OQ scales are heterogeneous and ill-defined, and they call into quesiton the intended use of the instrument, particulary the scale scores, in clinical settings.

A methodologically rigorous comparison of the intended structure of the OQ-45 to plausible alternative structural models of the items, across time points, could guide

researchers and clinicians on how to best use and interpret OQ-45 data. Proposals for the improvement of interpretation of scores will be made, with an eye to maintaining the ways the OQ-45 currently meets the diverse needs of clinics.

The OQ-45 was designed to achieve a laudable goal – to efficiently and broadly assess overall functioning, in a way that is sensitive to change. It may function in many ways like a self-report measure of the Global Assessment of Functioning (GAF), the 5th axis of the *DSM*-IV-TR, American Psychological Association, 2000). Such a measure has the potential to objectively track the progress of many clients in a busy clinic, where limited resources may contraindicate a full diagnosis and assessment of functioning at every session. Such a full diagnosis on the part of the therapist may also be contraindicated by the model of therapy, such as in family therapy, where it might be inappropriate as well as impractical to assess the overall functioning of each member of the family within the session. The OQ-45 also serves as a screening tool for clinicians by asking about a variety of low-base rate, but important issues, such as suicidality and work violence. The practical need for a measure which effectively accomplishes what the OQ-45 strives to accomplish is illustrated by its wide adoption in clinics around the nation and its translation into half a dozen other languages over the last decade. Thus, the current project assesses the OQ-45 critically, but with appreciation for its merits and an understanding that it meets practical needs that have not heretofore been better met.

Personality Attributes in Relation to Clinical Presentation and Treatment

The Big Five (Extraversion, Neuroticism vs. Emotional Stability, Conscientiousness, Agreeableness, Openness/Intellect) has become, among researchers, the standard model of normal adult personality attribute variation (John & Srivastava, 1999) and it appears more frequently than other personality attribute models in current psychopathology literature. The Big Six (which adds a dimension of Honesty/Propriety and slightly alters the content of Agreeableness) has been proposed as an update to the Big Five, but is a similar model in most respects (Thalmayer et al., 2011). My review of relations between personality attributes and clinical presentation and outcome is organized by these models.

Self-report scores on Big Five personality trait dimensions in adulthood have been reliably associated with clinical presentation and disorders in many studies. This is unsurprising, as scores on Big Five (and Big Six) dimensions relate to virtually all important life outcomes, including longevity, health, divorce, friendships, arrests, and academic, professional, and creative success, etc. (Nettle, 2007; Funder, 2010). The specific ways that personality attributes relate to disorders, and the ways that such relations are differentially observable when using slightly different models of personality attribute structure and of mental disorder nosology, will be reviewed and compared.

Less research has addressed how personality traits moderate treatment in psychotherapy, but this question is of particular interest and the literature pertaining to it will be carefully reviewed. Considering that typically only about half of clients reliably improve in symptoms after a course of therapy (Westen, Novotny, & Thompson-Brenner, 2004), and that about half of patients who begin therapy drop out prematurely (Wierzbicki & Pekarik, 1993), making sense of how measureable individual differences between clients relate to such failures in treatment could ultimately aid the field in improving services and reducing suffering in a greater proportion of those who seek therapy. Knowing how individual differences in demographics relate to typical presentation and outcome is considered an important aspect of competency (Sue, 2006). Perhaps someday knowing how personality attributes moderate the success of psychotherapy will also be considered a core competency.

Big Five (and Big Six) Relations to Clinical Presentation

Clark (2005) proposes that both disorders and personality traits can most parsimoniously be conceptualized as manifestations of underlying temperament dimensions. Three temperament dimensions appear to be observable in early infancy. Negative Emotionality maps on to the adult dimension of Neuroticism, Positive Emotionality maps on to the adult dimension of Extraversion, and Disinhibition is analogous to a combination of low Agreeableness and low Conscientiousness. This framework posits the broad class of Externalizing disorders as stemming from Disinhibition, and the broad class of internalizing disorders as being related most closely to Negative Emotionality. (Externalizing tendencies are alternatively referred to as social maladaptation, and internalizing as emotional distress by Dishion and Stormchack

[2007].) Low Extraversion is specifically related to depression, and to a lesser extent social anxiety and schizophrenia (2005). I approach this topic making the assumption that personality attributes lead to vulnerability (predisposition) to certain disorders. While personality traits and disorder symptoms likely arise from the same basis of innate temperament (elaborated through developmental experiences) – Clark (2009) provides evidence that the elaboration of personality traits generally precedes the onset of disorder tendencies. (There are also other kinds of relations between traits and disorders, for example, traits that change after the experience of a disorder (scar), or that provide information about the likely prognosis (pathoplasty).

In exploring how personality attribute dimensions relate to clinical presentation, presentation can be defined in a variety of ways. The *Diagnostic and Statistical Manual* (DSM; APA, 2000) diagnoses are standard, detailed categories generally relied on in research studies. It is useful to discuss what personality traits relate to risk of meeting criteria for a specific disorder. However, the specificity of clinical diagnoses can be misleading. For one thing, people who meet criteria for one mental disorder often meet criteria for another as well – comorbidity of disorders is more the rule than the exception (Krueger & Markon, 2006). Additionally, different people may meet criteria, and thus 'have the disorder', with markedly different patterns of symptoms, making a category less specific than may seem to be implied. Further, empirical evidence in some cases does not support DSM categories (e.g. Coker & Widiger, 2005), but their use in professional settings and research focused on the category can lead to a kind of reification -- because there is a name and a body of literature, it must exist as a natural category. From a family systems perspective, DSM diagnoses are also problematic because they locate a problem within one individual, rather than taking familial and social context into account when seeking to identify and treat problems in well-being and functioning (Minuchin, 1974). Diagnosis is considered paramount in some clinical settings, and is typically relied on heavily in clinical trials for treatments, which isolate a specific disorder from other disorders. However, a critique of such studies is that they create internal validity at the expense of ecological validity – such a study can tell us that one treatment for a pure case of major depression works slightly better than another, but pure cases of major depression are actually rare in practice.

For these reasons, in addition to reviewing the literature establishing personality attribute relations to DSM disorders, I will also consider how attributes relate to other ways of conceptualizing clinical presentation. For example, a higher-order framework of disorder tendencies, externalizing and internalizing, has been proposed based on empirical evidence of consistent patterns of correlation and comorbidity between DSM disorders (Kreuger & Markon, 2006). As this framework focuses on vulnerability to a general type or broad category of difficulty in functioning, it would appear to lend itself well to relations with the general behavioral tendencies captured by personality attribute dimensions. I am also interested in considering clinical presentation from the perspective of general functioning, such as that captured by the DSM's Axis V Global Assessment of Functioning (GAF). This kind of parsimonious summary of well-being and overall functioning in social roles and interpersonal relationships may capture the basic information about extent of a client's difficulties, and how well they are improving or declining, that a busy clinic finds most useful. How such ratings might relate to personality attributes could therefore also be of interest to clinicians attempting to make efficient use of available information on clients.

Relations between Big Five/Big Six dimensions, DSM disorders, and internalizing and externalizing disorder tendencies. Many studies report correlations between Axis I or Axis II diagnoses and Big Five attribute dimensions. (Clark [2005] posits that the relation of personality traits to Axis II personality disorders should not be given special status in the study of personality and psychopathology because traits demonstrate equally strongly correlations with Axis I disorders.) Psychopathology is not an entirely separate domain from normal personality, the two realms share temperamental antecedents (Clark, 2005), and the domain of mental disorders draws heavily on specific aspects of normal personality traits. A map of how disorders and traits co-vary would not sort the DSM disorders evenly around the high and low ends of the five or six attribute domains. Instead the bulk of disorders would be seen to cluster around the high end of Neuroticism, with the remaining attribute domains often providing important secondary differentiation between the disorders.

Neuroticism vs. Emotional Stability/Resiliency (N). Nettle (2007) defines this trait dimension from an evolutionary perspective, as a measure of the responsiveness of

the negative emotions, a system designed to warn us of potential threats in our environment. He argues that natural selection has favored relatively responsive systems, as negative emotions are less detrimental to our reproductive fitness than is an early death. Thus, many of us experience more negative affectivity than is warranted by the actual threats in our lives. And for those at higher levels of this normally distributed trait, there is a risk of internalizing disorders, like anxiety and depression (2007). Caspi and colleagues provided further support for the view that N is fundamentally linked to an internalizing disorder tendency by demonstrating that the same brain region responsiveness (amygdala) linked to depression is also linked to increased scores on self-report measures of N (2003).

Claridge and Davis (2001) complain that N is so highly and consistently related to psychopathology as to lack discriminant power. Scores on N predict pain, psychosomatic complaints, cold symptoms, nightmares, PMS, eating disorders, depression, suicidality, drug abuse, obsessive compulsive disorder (OCD), and schizophrenia. In their studies, while highly correlated with symptoms, N did not explain unique variance when a more specific scale was introduced to account for symptoms (for example, nightmares with schizotypy; weight preoccupation with narcissism, borderline personality, and obsessive-compulsiveness). Some of the lack of specificity found by these authors may be due to their reliance on Eysenck's EPQ-R (1991) to measure N, which, like the NEO-PI-R (Costa & McCrae, 1989), combines impulsiveness, anger and anxiety within N. Other authors have clarified that anxiety and impulsivity can be better viewed as separate aspects of temperament (Clark, 2005; Rothbart, 2007), brain activation (Carver, 2005), and personality (Nettle, 2007), with impulsivity defining the low end of the Conscientiousness dimension (and to some extent, high Extraversion).

Despite this limitation, Claridge and Davis (2001) make an interesting case for using N as a moderator variable when studying other constructs related to disorders. For example, in the relation between sedation-response and extraversion, N was associated with both extremes of drug tolerance (high resistance to sedation and high arousability were related to introversion with high N; low resistance to sedation/low arousability was related to extraversion with high N.) This was also true of three aspects of body image in women. A measure of narcissism was found to have a positive correlation with body

esteem for women with low to average levels of N, whereas at high N, narcissism had no positive effect on body image. Likewise a measure of masculinity in women was related to higher body satisfaction, but only at low to average levels of N. And perfectionism, which is generally thought to decrease body esteem, did not have this effect for women low in N (Claridge & Davis, 2001). The authors conclude that N may function as a moderator because negative affectivity "potentiates negative elements in co-existing features of the individual, transforming them from adaptive, to unhealthy, behaviors" (2001, p. 395).

In Clark, Watson, and Mineka's tripartite model of mood disorders (1994), N is a general predictor of depression and anxiety, but it is moderated by E (discussed below) and a dimension of reactivity specific to anxiety. They cite evidence for a variety of interactions between N and depression diagnoses. For example, N scores appear to be elevated after a depression (congruent with a "scar" model of the relation between traits and psychopathology) and high N score leads to slower recovery and more likely recurrence of depression, irregardless of the severity of a depression episode (congruent with a "pathoplasty" model). Score on N also appears to have state variation -- lower scores are seen for those in remission from depression (congruent with a continuity model).

Can Neuroticism itself be treated? Personality attributes are known to be largely, but not entirely stable, and Emotional Stability typically increases during adulthood (Roberts, Walton, & Viechtbauer, 2006; Srivastava et al., 2003). Knowing that normative changes are to be expected may allow for fruitful explorations of how to activate or accelerate such change. De Fruyt and colleagues (2006) note that while personality continuity is a precondition for using scores to predict disorders, studies often show decreases in N scores after treatment, and it appears to be the most changeable of the Big Five after therapeutic interventions. Selective serotonin reuptake inhibitors and noradrenergic antidepressant agents appear to alter both depressive states and personality attributes. A group of patients treated for depression over a six month period, saw increases in Emotional Stability, E, O, A, and C scores, but patients' Emotional Stability was still lower than the population mean (De Fruyt et al., 2006). A significant component of stability should be expected for this tendency to reactivity in the negative

emotions, even when successful interventions draw scores closer to the mean. Dramatic reversals may be unrealistic, but even small changes may indicate important gains in well-being.

Conscientiousness vs. Disinhibition. A dimension of constraint and social self-regulation versus undercontrolled behavior appears relatively consistently across personality trait structural models in diverse cultures (John & Srivastava, 1999; Saucier, 2009; Ashton & Lee, 2010), and across various temperament and biobehavioral models (Carver, 2005; Clark, 2005: Rothbart, 2007). Nettle (2007) summarizes brain functioning correlates with self-report inventories to define this dimension as the strength of response inhibition in the dorsolateral prefrontal cortex. As such, it appears to be implicated in all disorders captured under the higher order dimensions of Externalizing. In models of temperament, the term Disinhibition is more commonly used, and it is sometimes described as including content also related to adult Agreeableness. However, in adult personality, Conscientiousness is associated directly with response inhibition and retains predictive power for disorders, while Agreeableness generally does not (Nettle, 2007). It may be that Conscientiousness is a more direct adult expression of the temperament domain related to effortful control.

Low scores on this dimension predict externalizing disorders in childhood and adulthood, including Attention Deficit Hyperactivity Disorder (ADHD; Clark, 2005). In adulthood, when associated with Negative Affectivity, low Conscientiousness predicts substance use and problem gambling (Slutske, Caspi, Moffitt, & Poulton, 2005). Extremely low Conscientiousness is associated Antisocial Personality Disorder, a syndrome characterized by disregard for the law and the rights of others, deceitfulness, impulsivity, irritability, aggressiveness, recklessness, irresponsibility, and lack of remorse (APA, 2000). Some conceptualizations stipulate that the high Disinhibition characteristic of the disorder is associated with aspects of low N (a rarity in psychopathology relations to N) particularly low anxiety and fear (Coker & Widiger, 2005). ASPD is also associated with aspects of low Agreeableness (high arrogance and low empathy), and high Extraversion (glib charm and sensation seeking; Coker & Widiger, 2005).

Low Conscientiousness (C) is also implicated in Borderline Personality (BPD), a syndrome of intense fear of abandonment, feelings of emptiness, impulsivity, anger,

dichotomized thinking and lack of stability in relationships, identity and behavior (APA, 2000). This disorder is associated with extremely high N, as well as aspects of Agreeableness (low trust and compliance), and C (Lowe & Widiger, 2008). Researchers seem to agree that BPD arises from a toxic combination of biological vulnerability (a temperament high in N and Disinhibition), coupled with childhood adversity and ongoing problematic interpersonal relationships (Lowe & Widiger, 2008). While ASPD is more common in men, BPD is more common in women (APA, 2000).

High C can also be problematic when paired with high N. Obsessive-compulsive personality disorder (OCPD), a rigid fixation on rules, order and control, perfectionism, hoarding, and/or excessive devotion to work (APA, 2000), is characterized by very high levels of C (Samuel & Widiger, 2011) as well as aspects of Agreeableness (angry hostility), Extraversion (very low excitement seeking), and Openness (Lowe & Widiger, 2008). While OCPD is more common in men than women, a similar combination of high N and high Conscientious (particularly perfectionism) may be related to anorexia nervosa and other eating disorders, which are more common among women (APA, 2000), and which share a common familial liability with OCPD (Lilenfeld, Wonderlich, Riso, Crosby & Mitchell, 2006). Nettle, (2007, based on prevalence data reported by Coid et al., 2006) notes that while average levels of C appear to be equal for men and women, the disorders of extremely low and extremely high C are more prevalent in men than in women by a factor of 2:1.

Extraversion. Probably best conceptualized as responsiveness to available rewards in the environment (Nettle, 2007), this is an appetitive, motivational approach system, a biobehavioral dimension of "increased locomotor behavior, incentive motivation, sensitivity to reward signals, interest and alertness, euphoria, excitement and pleasure seeking on the one hand, versus low energy and activity levels, withdrawal, decreased cognitive capacity, anhedonia and depressed mood on the other" (p. 107; Clark et al. 1994). This dimension is highly related to the temperament dimension of Positive Affectivity (Clark, 2005). Extraversion may act as a protection against internalizing disorders, but in some manifestations it may confer risk for externalizing disorders.

When associated with high N, low Extraversion predicts melancholy depression. It is correlated with measures of depressive mood and cognitions, but not with anxiety

(Clark, 2005), thus it seems to relate to the bifurcation of internalizing disorders into distress and fear (Kreuger & Markon, 2006). However, it may have some relations with social phobia, an anxiety disorder (Clark, 2005). Dysregulation of the positive emotions and reward approach system is also associated with bipolar disorder, and with the diurnal mood variations observed in depression (Clark et al., 1994).

Low Extraversion (E) also appears to be related to schizophrenia and schizophrenia spectrum disorders (Clark, 2005), including the DSM Cluster A "odd or eccentric" Axis II disorders (Paranoid, Schizoid, and Schizotypal Personality Disorders), conceptually grouped by social withdrawal, all of which appear to be more common in relatives of individuals with schizophrenia (Lowe & Widiger, 2008).

High E protects against depression and leads to quicker recovery from mood disorders, even in severe cases (Clark, 2005). In many senses, high E may define those who are particularly psychologically healthy (Clark, 2005). However, some aspects of high extraversion may be present in personality disorders, including BPD and ASPD as discussed above, as well as in Narcissistic and Histrionic personality disorders. These latter two disorders display particularly high levels of E, with flamboyant, assertive, gregarious behavior coupled with high N (though not as high as in BPD) and aspects of low Agreeableness (low modesty and altruism; Coker & Widiger, 2005). The attractive qualities of E, even when coupled with problematic behavior, is illustrated by a study that found that dorm dwellers and military recruits rated peers with histrionic and narcissistic traits highly as people they would like to get to know (Oltmanns & Turkheimer, 2009). This speaks to how such disorders may be exacerbated and maintained – a style that hides insecurities and garners attention in a new social setting may reap short-term rewards, though difficulties are likely to develop over time, and those with histrionic and narcissistic traits were more likely to have been dismissed prematurely from the military a year later (Oltmanns & Turkheimer, 2009).

Agreeableness. Big Five Agreeableness (A) is associated with positive experiences and life and outcomes, such as a sense of humor, religious involvement, a healthy heart, peer acceptance, and dating satisfaction (summarized by Funder, 2010), and nonclinical disadvantages like diminished creative achievement, wealth and status, and less success in business in terms of promotion and leadership (Nettle, 2007). There

is some divergence of opinion in the definition of Agreeableness. The core of this dimension across diverse lexical studies emphasizes gentleness of behavior and not getting in the way of others (recurrent content include being helpful, peaceful, and tolerant versus being domineering, aggressive, bossy, or egocentric; Saucier, Thalmayer, & Bel-Bahar, 2012). The NEO-PI-R, used predominantly in recent studies of psychopathology, is a more broad scale of interpersonal virtues or "niceness" (facets include trust, straightforwardness, altruism, compliance, modesty, and tendermindedness.)

Trull & Sher (1994) claimed that a NEO profile of high N and O, and low E, A, and C (basically, the undesirable poles of all dimensions) characterized all those with Axis I pathology in a sample of young adults, and that a profile of low N, C, and A, and high E characterized the substance users who did not have comorbid depression (externalizers). Lowe and Widiger (2008) found that NEO aspects of Agreeableness differentiated dependent personality disorder from Avoidant PD (for what such a distinction is worth, given the high comorbidity between the two) and also observed associations between specific facets of Agreeableness and BPD, ASPD, Histrionic PD and Narcissistic PD as described above. Aside from these studies, and a negative correlation with criminal behavior (externalizing tendencies; Ozer & Benet-Martinez, 2006) there are not strong or consistent relations between A and psychological disorders. Lowe and Widiger (2008) point out that the NEO-PI-R may be especially limited for the purpose, because it does not assess for maladaptive levels of Agreeableness. Big Six conceptualizations of Agreeableness (and that of the Big Five Inventory) may be better suited to distinguishing between disorders because impatience and anger are located on the negative pole (Thalmayer et al., 2011), which should allow for stronger associations between A and externalizing disorder tendencies.

Openness/Intellect. Some authors (e.g. Saucier, 1995) have noted that the dimension of Openness (O) has not shown consistent relations to psychopathology. While Soldz & Vaillant (1999) found measures of Openness (similar to NEO Openness) administered in college to be predictive of psychiatric visits (as well as creativity and less conservative attitudes) over a subsequent 45-year period, psychiatric visits is an indirect stand-in for psychological problems. A plausible alternative explanation is that high

scorers in this domain may be particularly interested in self-development and personal growth (DeYoung, Peterson & Higgins, 2005; Thalmayer & Saucier, 2011), leading to an adaptive use of psychological services.

Lowe and Widiger (2008) argue that the ambiguous results to date relating Openness to psychological disorders are due to limitations of the NEO-PI-R conceptualization of Openness, which does not define the dimension strongly in terms of unconventionality, or assess for extreme, potentially maladaptive ends of facets. They argue that a measure with such characteristics could make scores on Openness function as indicators of Schizophrenia spectrum personality disorders. This opinion would appear to be shared by Nettle (2007), who notes the ambiguity of content in this factor across inventories, but cites associations between high scores, creativity and psychotic experiences and disorders, and ultimately defines the dimension as indexing breadth of mental associations. DeYoung, Peterson, and Higgins (2005) specifically characterize Openness/Intellect as "motivated cognitive flexibility" (p. 850), or cognitive exploration, quite possibly modulated by dopamine. Recent evidence supports the relation between creativity and dopamine availability (specifically in the thalamus) and the relation between creativity, divergent thinking and the genetic liability for schizophrenia spectrum and bipolar disorders (de Manzano, Cervenka, Karabanov, Farde, & Ullén, 2010). Brain networks relevant to divergent thinking overlap with regions and networks affected in schizophrenia and bipolar disorder (de Manzano et al., 2010). Thus, increased attention to the specific content of this dimension could lead to a dimension that more consistently demonstrates association with specific mental disorder tendencies.

Honesty/Propriety. Six-factor models of traits have been developed from work with lexical studies in diverse languages, in which comparable six-factor solutions have emerged more reliably than five-factor solutions (Lee & Ashton, 2004; Saucier, 2009). Altering the content of Agreeableness (to focus on immunity to anger, undemandingness, forgiveness, patience, trustfulness) and adding a dimension of Honesty and Propriety with content related to ethical behavior (honesty, humility, and integrity, non-instrumentality, norm-inviolacy and safety-mindedness) allows for a structural model of attributes that is more robust across cultures and across variable selection strategies (Saucier, 2009). Such alterations increase explanatory power over that of the Big Five by

allowing for a more elaborated model of important individual differences in behavioral tendencies (Lee & Ashton, 2004; Saucier, 2009; Thalmayer et al., 2011).

The dimension of Honesty/Propriety is hypothesized to be useful for predicting criminality and externalizing disorders. The content of this domain (some of which may relate to facets of Conscientiousness or Agreeableness in some inventories, but most of which has been left out of the Big Five) relates to regulating behavior by rules and norms governing social relations. While Conscientiousness predicts to punctuality and organization, Honesty/Propriety is conceptually related to degree of attention to normative expectations set out in a social context, and to following through with commitments. For example, it has demonstrated superior predictive power as compared to C and other personality dimensions, for grade point average in a sample of college students (where the range of C is likely to be restricted; Thalmayer et al., 2011).

Research relating this dimension to disorder tendencies is more limited than research using Big Five inventories. However, findings to date support this as a domain relevant to externalizing disorder tendencies. The dimension of Honesty/Propriety was shown to correlate more highly than other Big Five or Big Six dimensions with compulsive drinking, risk taking, and lawbreaking behaviors (Saucier, 2009). The dimension (termed Honesty/Humility in the HEXACO inventory) has also demonstrated strong negative correlations with measures of the 'dark triad': psychopathy, Machiavellianism and narcissism (Lee &Ashton, 2005). It would be logical if this dimensions allowed for a better articulated distinction between aggressive behaviors, predicting instrumental aggression more strongly than A, which should relate directly to (low) reactive aggression.

Big Five/Big Six relation to overall functioning. Does an individual's pattern of scores on a Big Five or Big Six measure provide an assessment of general psychological functioning? Are some scores better than others? Some researchers suggest that fluctuating pressures on natural selection has led to the persistence of variation in the trait dimensions across time because there is a balance of benefits and drawbacks spread across high and low levels of personality attribute dimensions, especially if variations in cultural demands, environmental conditions, and social structures across time are taken into account (Nettle, 2007; Saucier, 1994). Hogan (1983), who identified Conformity and

Sociability as the common denominators of circumplex models of personality that account for individual differences in vocational roles, similarly argues that having members who represent a spectrum of values on these dimensions (maintainers of the culture vs. creative innovators; and those who work to keep the keep the group together vs. those who work alone to develop tools and technology) is evolutionarily adaptive for groups. In his model, other personality traits and dimensions are seen as relevant for describing how well individuals perform within their social roles (Hogan, 1983).

From a clinician's perspective, in contrast to the highly evaluative 'personality' measures developed from the psychodynamic perspective (such as the MMPI, the Personality Assessment Inventory, and the quality of object relations) the Big Five come across as refreshingly descriptive and easy to discuss with clients, with only the term Neuroticism (which is termed negative emotionality in NEO feedback materials) having an obviously negative meaning (Coleman, 2006). However, in the privacy and context of a therapy session, a digestable amount of evaluation is likely appropriate, and Coleman (2006) explicitly discusses the advantages of higher levels of Extraversion and Agreeableness, and lower levels of Neuroticism, and identifies these attributes as targets of therapeutic interventions.

I would argue that the names given to the Big Five dimensions make desired scores relatively explicit. If the Big Five were truly only descriptive, and it were equally advantageous to be at any level on the dimensions, then the evaluative tone of the labels should be questioned. This evaluative tone likely accounts for the lack of adoption of the Big Five model of personality structure outside research settings, and, instead, the continued popularity of the Myers Briggs Type Inventory, in which all personality types are presented as equally desirable, in occupational and social settings.

On balance, however, the evidence suggests that there are relatively more advantageous positions on the scales of trait dimensions, and the evaluative tone of the Big Five dimensions labels reasonably accurately reflects the nature of those dimensions. For example, while the complete lack of anxiety may be associated with psychopathy, only moderately higher than average levels of Neuroticism predict a host of physical, psychological, and environmental difficulties (Nettle, 2007; Funder, 2010). Likewise, while abnormally high Conscientiousness is related to a personality disorder and an

eating disorder when coupled with high Neuroticism, above average scores on this dimension are generally related to successful functioning in society.

Languages the world over include words to describe psychological differences between persons, and the words that appear most commonly are quite evaluative in nature (Saucier et al., 2013). It may be that the language of personality that has developed in every known human language, has developed in large part to evaluate general functioning in others, and to communicate desired behavioral attributes while socializing children. The Big Five, which were identified initially in studies of the natural language used to describe persons (John & Srivastava, 1999), may have been drawn from just such a pool of naturally evaluative terms.

Thus, there might be a way to create or to score a personality inventory that would allow for a general measurement of functioning. Above average Emotional Stability, Conscientiousness, Agreeableness, Extraversion, Honesty/Propriety and perhaps Openness would constitute a favorable personality profile, and would be expected to predict higher functioning. It would be interesting to explore the extent to which such a favorable personality profile correlated with direct measures of psychological functioning, such as the OQ-45 total score, or the DSM's global assessment of functioning (GAF). Such measures designed to assess functioning in a clinical population would naturally be expected to cover a broader range of pathological functioning, and thus might pick up where a rating of general personality functioning leaves off, while still evidencing high correlations. (In a small sample [N=63] GAF was associated with Big Five Neuroticism [r = .52; Coleman, 2006]. However, no other research comparing personality measures to the GAF have been identified; nor have any studies comparing the GAF to the OQ-45 total score been identified.)

Summary. Personality attribute dimensions relate to psychological problems in terms of DSM disorders, overarching disorder tendencies categories (internalizing and externalizing), and very likely, general functioning. Big Five attribute scores may tell us more than a DSM diagnosis, because they are broader and more stable than diagnoses (Clark, 2009). Most psychological disorders, especially those encompassed by the broad heading of internalizing disorders, are associated with increased negative emotionality (Clark, 2005). (In fact, this relation is so core to the content of this attribute dimension

that it is labeled Resiliency vs. Internalizing Negative Emotionality in the recently developed QB6 measures of the Big Six; Thalmayer et al., 2011). The domain of Conscientiousness (or sometimes a higher order construct of Disinhibition, which also includes Agreeableness) has been strongly negatively associated with externalizing behavior tendencies (Clark, 2005). This relation is hypothesized to be equally strong for the Big Six dimension of Honesty/Propriety, which has demonstrated relations to narcissism, psychopathy, and Machiavellianism (Lee & Ashton, 2005), and to problematic drinking and risk taking (Saucier, 2009). The two externalizing-related dimensions should provide for predictive distinctions, likely along the lines of reactive versus instrumental aggressive and rule breaking behaviors. Extraversion, the domain of reward motivation and positive emotions, may increase risk of externalizing disorders, but also works as a protector against internalizing type problems. Levels of Extraversion are likely associated with the bifurcation of internalizing disorders (Krueger & Markon, 2006) into distress (low Extraversion) and fear (less association with Extraversion). Agreeableness is associated to some extent with externalizing tendencies, especially when the low end of the scale is defined with hostility and impatience content. Openness may be related to perceptual abnormalities in psychotic experiences that are uncommon and fall outside of the internalizing and externalizing disorder tendency domains.

Personality dimensions likely also relate to measures of general functioning, such as the OQ total score and the DSM Axis V GAF (above average Emotional Stability, Conscientiousness, Agreeableness, Extraversion, Honesty/Propriety and perhaps Openness would be expected to predict higher functioning). There may also be a way to assess general functioning with a one-factor measure of personality.

Big Five/Big Six Relations to Clinical Treatment and Outcomes

In summarizing the lack of research relating normal personality variation to clinical outcomes, Coleman (2006) notes, “Personality is such a basic element of human experience that it often goes unnoticed, like the air we breathe. Just as we notice polluted air, we tend to notice personality when it is interpersonally troublesome.” (pp. 83-84) This reflection captures the experience of a personality researcher entering clinical or social work arenas, where the term ‘personality’ is associated almost exclusively with personality disorders or the dimensions of neuroses on the MMPI. The rich knowledge

base of personality psychology appears to be sorely underused in applied settings. Considering the moderate rates of efficacy for virtually all forms of psychotherapy (Westen & Bradley, 2005) and high drop out rates (Wierzbicki & Pekarik, 1993), there is a great deal of variation left to account for. This neglect of well-validated and easy-to-measure individual psychological differences may someday be viewed as a gross oversight.

Miller (1991) provides an intriguing initial map for how Big Five trait scores may relate to responsiveness and outcomes in therapy. Drawing on a sample of 119 clients seen in his private practice over a two-year period (101 therapy seeking and 18 seen in the context of someone else's therapy), Miller makes a case for the utility of personality measurement in appropriately tailoring treatments and expectations.

Unsurprisingly, high scores on Neuroticism are to be expected in those who seek treatment (Miller, 1991). Knowing a patient's standing with regard to this dimension may aid in contextualizing the presenting complaint and in setting appropriate expectations for post- treatment well-being and functioning. Neuroticism has been associated with less treatment success for depression (Mulder, 2002; Quilty et al., 2008). However the long term misery associated with N may function a motivator for therapy and making life changes (Miller, 1991).

Among the Big Five (or Big Six) dimensions, Neuroticism may be the best predictor of psychological distress and therapy usage, but Conscientiousness (C) might best predict therapy success (Miller, 1991). Psychotherapy may have basically been designed for those higher in C, who have the self-discipline to diligently follow the instructions of interventions to make changes in their lives. Miller (1991) identifies C scores under 40 on the NEO-PI-R as so low that most types of therapy may be ineffective. Later studies have indeed associated higher pre-treatment C with more successful response to treatment (Quilty et al., 2008), even accounting for life stress and severity (Anderson & McLean, 1997).

The role of scores on the dimension of Honesty/Propriety (H/P) in treatment preferences and outcomes have not been explored, but would logically be as crucial as C. A tendency to avoid taking advantage of others and to aim for straightforwardness and integrity with others may provide a necessary foundation for the challenges of therapy.

Individuals high on H/P could be expected to take responsibility for personal failings, and to have some motivation for self-improvement. As discussed above, individuals with especially low scores on Honesty/Propriety may be at risk of externalizing disorders, especially the 'dark triad" of Machiavellian, narcissistic and psychopathic tendencies (Lee & Ashton, 2005). Such individuals would likely avoid therapy unless mandated, or unless a secondary gain provided motivation.

Determining if (and what) minimum levels of Conscientiousness and Honesty/Propriety are necessary for psychotherapy to work as intended may meaningfully improve outcome research and services. If persons with very low scores are unable to benefit from most forms of therapy (as observed by Miller [1991]), including them in studies of therapy outcome and efficacy may obscure the actual effects of treatments. Instead, alternative or additional interventions (such as motivational interviewing) should be explored and tested. Offering people treatments they cannot benefit from may cause harm. For example, there is some evidence that counseling for domestic violence offenders has little effect on their likelihood of reengaging in violence, but it does influence partners to go back to the (unsafe) relationship, and diverts funds from resources for battered women and children (Feder & Dugan, 2004).

Quilty and colleagues (2008) demonstrated that Extraversion and Openness increased response to psychopharmacological and psychotherapy treatments for depression. And Coleman (2006) provided evidence that Agreeableness, Openness, Extraversion, and Conscientiousness contributed to the formation of a positive working alliance in community mental health center clients, a common factor in therapy known to correlate with beneficial therapeutic outcomes. While Miller (1991) also associates scores on those dimensions with the easier formation of a working alliance, his overall view of the role of Agreeableness, Openness, and Extraversion in therapy is mixed. He suggests that while persons high in Agreeableness may be warm and gratifying to work with, a need to please the therapist may make it hard for such clients to assert themselves and think clearly about their own best interests. It is also plausible to hypothesize that clients high in Agreeableness may stay in therapy longer than needed, attending sessions compliantly whether or not they are getting better.

Likewise Miller (1991) suggests that therapists are likely to view individuals higher in Extraversion as healthier, because they are cheerful and forthcoming, but that such characteristics might also mask the misery that brought an individual to therapy. He also suggests that therapists of some orientations may prefer high Extraversion clients because they more fluently express themselves and disclose personal information, and that clients low in Extraversion may prefer more structured approaches, such as cognitive therapy (Miller, 1991).

Scores on Openness have been associated with successful treatment of depression in one study (Quilty et al., 2008), but the association is not well explored. Interestingly, Miller (1991) suggests that scores on Openness may have special significance in terms of the choice of therapy model. In fact, differences in Openness may help elucidate long running theoretical differences in treatment models. Miller (1991) suggests that high Openness is especially attractive to therapists of some orientations:

> The existentialist-humanist school regards O as an end in itself (e.g., Rogers, 1961). The psychoanalytic school implicitly demands the ability to fantasize and symbolize; failure to do so is regarded as resistance, which the therapist is expected to help the client overcome. Thus, a high N, high O client who is able to fulfill the demands of the treatment model is going to seem healthier to most therapists than a high N, low O client who, despite considerable distress, free associates about cleaning the garage and the neighbor's barking dog. Therapists who understand the O domain will be less likely to make potentially harmful value judgments about such clients or themselves. (Miller, 1991, p. 425)

Openness is positively correlated with fluid intelligence (Moutafi, Furnham, & Crump, 2006) and creativity (McCrae, 1987). This might help explain a predilection of early psychological leaders of the humanistic and existentialist schools, for this trait. Such thinkers may conceivably have been biased by their high valuation of their own best qualities when they equated such characteristics with mental health. For example, Rollo May (1967) articulated a conception of psychological maturity that he believed would account for a decrease in neurotic anxiety. He stated that a shift in values, from immature (emphasizing physical gratification and survival) to mature (transcending the immediate situation and in-group) would counter such anxiety. This theory was recently tested by assessing whether higher scores on a scale designed to capture May's Mature Values would lead to increased subjective well-being, decreased psychopathology, and positive

changes in Big Five personality trait scores in a longitudinal sample of mature adults (Thalmayer & Saucier, 2011). In fact, the scale was not significantly associated with changes in well-being or depression (although there were negative associations with obsessive compulsive symptoms, narcissism, materialism and Machiavellianism, among other correlates). Instead, it most strongly predicted an increase in Openness (Thalmayer & Saucier, 2011).

The possibility exists that therapists may prefer clients high in Openness, without this dimension having a strong relation to mental well-being or potential for response to treatment. On balance however, the core characteristics of Openness/Intellect (intellectual curiosity and imagination) would appear to contribute to the process of psychotherapy in many treatment settings.

Summary. Big Five/Big Six relations with the outcome of clinical treatments are not well known at this time. However, there is every reason to believe that these attribute dimension scores will be associated with therapy-relevant preferences, motivation, and behavior, and that taking attribute dimensions into account could help to explain variance in therapy efficacy and drop out rates.

Initial evidence suggests that those high in N make up the majority of persons who need and seek psychological services. Otherwise, a more positive personality profile, as discussed above (higher A, C, E, and O), may be beneficial for psychotherapy, in much the same way that such a profile leads to benefits across life domains. Extraversion and Agreeableness, in particular, have in some cases been associated with better therapeutic outcomes, and in other cases have shown no significant relation, but lower scores on these dimensions have in no cases been shown to benefit therapy.

More significantly, minimum levels of Conscientiousness and Honesty/Propriety may be necessary for psychotherapy to work as intended. A score cut off on both H/P and C for engaging in therapy may ultimately prove to be appropriate.

Despite some evidence for an association between Openness/Intellect and psychotic experiences, there is no evidence that high O is detrimental to therapy, and some evidence that it aids the process (Coleman, 2006; Quilty et al., 2008). It is suggested that O may be particularly associated with the values inherent to the types of therapy most commonly available in the last century – existential, humanistic, and

psychoanalytic, and that individuals with low O may prefer more structured contemporary treatments, such as cognitive therapy (Miller, 1991).

Some Goals for Research

It is important to understand how client characteristics play a role in therapy use and outcome. From the beginning, attributes likely play a role in commitment to the process of therapy. At community mental health clinics, about half of clients who begin therapy drop out prematurely (Wierzbiecki & Pekarik, 1993). At a typical university family therapy clinic, about 30% of all individual, couple and family clients attend less than three sessions (Masi, Miller, & Olson, 2003). It would be unrealistic to expect all clients who attend an intake session at a clinic to stay for any particular amount of time -- it is easy to imagine that for some, therapy is not what they pictured, perhaps because an alliance is not quickly formed with the therapist (Egan, 2005). For other clients, spontaneous improvements in social support or life circumstances may make the need for therapy less pressing (Egan, 2005). Clinicians and researchers, however, should continue to explore how services can more effectively be administered to diverse clients in order to increase efficiency in mental health services (Masi et al., 2003).

Some associations between client characteristics and drop out have been established. While age and sex are not believed to be reliable predictors of premature termination, other client characteristics (lower SES, more severe symptoms, expectations, beliefs and attitude toward therapy, readiness for change) have been (Egan, 2005, provides a review). Normal-range personality attributes, however, have not been adequately explored as moderators of therapy usage and outcome success, although there is every reason to believe they function as such.

Previous studies have focused on traits in the treatment of specific disorders at specialized clinics. Building on these findings to better elucidate how personality attributes moderate the success of psychotherapy in a more general, community clinic could provide new insights into how personality attributes inhibit or aid treatment. Such knowledge could lead to improvements in services offered to clients, at low cost and minimal client effort. Such a setting provides for more variation in presenting complaint and level of function in clients, and could allow researchers to observe how the effect of traits varies when comparing individual to couple or family therapy treatment.

Measuring normal range traits could potentially give therapists a context for better understanding a client as a whole, and for identifying strengths that may contribute to treatment success. Such measurement might also reveal 'selection effects' – specific therapies or practitioners may select clientele based on personality profiles desirable to certain modalities. Ideally, treatments could be modified for individuals, taking strengths and challenges associated with low and high levels of traits into account when planning treatment. Alternatively, a complementary set of therapies that engage the full range of personality styles effectively can and should be identified.

CHAPTER III
METHODS

Study 1

In the first study, the intended structure of OQ-45 items was compared to alternative models and to a baseline "total score" model in separate portions of a large dataset using EFA and CFA. It is hypothesized that patterns of response on the OQ-45 may be driven more by basic, robust domains of individual psychological differences than by distinctions between the intended subscale domains. An alternative hypothesis would be that more specific problem areas than the intended subscales underlie scores on the OQ-45. Groupings might relate to internalizing/externalizing tendencies, or to more specific areas, e.g., substance use, depression, anxiety, conflict, etc. The alternative models were created by arranging the existing OQ-45 items into theoretically driven, subscale groupings using (see Appendix A for details). These included two personality models (a three-factor model with Negative Affectivity, Positive Affectivity, and Disinhibition scales, and a four-factor model adding Affiliation) and two specific problem models (a two-factor internalizing/externalizing model, and a seven-factor model). Refined, preferred models are compared in samples of responses from later therapy sessions.

Participants

Two thousand one hundred clients attended at least one session at the Couples and Family Therapy (CFT) clinic between January, 2006 and September, 2011 and completed the OQ on at least one occasion (this excluded most clients under age of 18, who more often completed a youth version of the questionnaire). The clients in this sample were 57% female, 85% white, and had an average age of 34 years (SD = 10.5). Half of clients came from households with less than $25,000 per year in income. About half attended the clinic for individual therapy sessions, and half for couple or family therapy sessions.

Materials and Procedure

All items of the Outcome Questionnaire-45 version 2 are answered on a 5-point Likert-type scale ranging from 'never' to 'almost always'. Nine items are reverse scored. The questionnaire was typically administered at the CFT clinic electronically, by PDA, before a client's intake, third, fifth, and tenth sessions, and every ten sessions thereafter.

In some cases a paper version was administered, and in some cases the OQ was administered prior to other sessions.

Analyses

The structure of the OQ-45 was first explored using exploratory factor analysis (EFA; oblique rotation) on a random third of the time-1 dataset. Factor scores for two-, three-, four-, and seven-factor solutions, plus solutions with increasing numbers of factors, until interpretability was lost (up to 10), were compared to scale scores for the *a priori* models using Pearson correlation. The empirically derived factors were matched, based on dominant content, to the *a priori* model factors. The correlations of the matched factors for each model with the appropriate empirical factor solution were averaged using z score transformations, in order to provide an approximate representation for the degree of fit between the hypothesized and observed models.

Secondly, the structural models and a baseline total score model were compared using CFA in the second random third of the dataset. All models were fitted in Mplus 7 using full information maximum-likelihood estimation. Because the data did not meet the assumption of multivariate normality (Mardia's coefficient of multivariate skew = 27,673, $p < .01$; multivariate kurtosis = 49, $p < .01$), robust maximum likelihood estimation (MLR) was used and adjusted chi-square values are reported (as in Bludworth et al., 2010; Coco et al., 2008). Modifications were made to alternative models, and the preferred models were tested in the last random third of the dataset. Preferred models were then tested again in time-5 and time-10 datasets.

In addition to adjusted chi-square, a mix of indices that evaluate different aspects of fit were examined, following the recommendations of Hu and Bentler (1999) and others. These included: One comparative fit index (CFI) comparing variance explained to a null model, and another (Tucker-Lewis; TLI_1) comparing improvement in fit over the baseline total score model; the standardized root mean square residual (SRMR), a measure of absolute fit; the root mean square error of approximation (RMSEA), which adjusts for parsimony; and Akaikes Information Criterion (AIC) and Bayesian Information Criterion (BIC) for comparison of non-nested models, taking parsimony into account. Parameter estimates are another crucial aspect of fit (Marsh et al., 2005). Kline (2011) states that an effective model should explain the majority of the variance in each

indicator. A count of how many items had model R^2 values over .5 is thus contrasted with a count of how many had R^2 values under .1, with the caveat that this count will naturally increase with the number of factors extracted. This count may be most useful in comparing models with equal numbers of factors.

Hu and Bentler (1999) have influentially suggested that a good-fitting model should meet several criteria, e.g., CFI > .95, RMSEA < .06, and SRMR < .08. Previous confirmatory studies using the OQ-45 suggest that such standards are not likely to be met. Research on the use of CFA in personality inventories also calls into question the likelihood that any model of responses on a measure like the OQ-45 could meet traditional standards. Kline (2011) points out that CFA is most effective where items can be parceled, rather than with item-level inventories. Even in longer inventories that allow for parceling, Hopwood and Donnellan (2010) demonstrate that well-validated multidimensional personality inventories routinely fail to achieve adequate fit per current standards. The OQ-45 may be similar to a personality inventory in capturing constructs that are both meaningfully distinct and meaningfully interrelated, such that an independent clusters model cannot provide good fit. Bludworth et al. (2010) adjusted *a priori* standards for the OQ-45 per Marsh et al. (2005), by emphasizing RMSEA and SMSR (cutoffs .08), and de-emphasizing incremental fit. They did not anticipate significant chi-square due to model complexity and sample size. In the current study, fit indices were assessed both in terms of general standards and in comparison to indices previously reported for the OQ-45.

Study 2

Study 2 allowed for the exploration of relations between OQ-45 items and Big Six personality dimensions, and for testing the structural models identified in Study 1 in a non-clinical sample. Resiliency, Extraversion, Honesty/Propriety and Conscientiousness, in that order, were hypothesized to correlate most highly with OQ-45 total score. The overlap and convergence of the content measured by the two questionnaires will be assessed and OQ-45 items suited to measuring personality attributes will be identified.

Participants

Undergraduate students in introductory psychology and linguistics courses (N = 589) completed surveys in Fall 2011. The sample was 64% female and 75% Caucasian

(10% “other”, 9% Asian, 5% African American, 2% Native American or Native Hawaiian). The average age of the sample was 19.5 (SD = 2.2).

Materials

The Outcome Questionnaire-45 version 2, as described above, and the 48-item Questionnaire Big Six Scale (48-QB6; Thalmayer et al., 2011). The 48-item Questionnaire Big Six Scale (48-QB6) was proposed to improve on the Big Five in terms of cross-cultural replicability and relation to important life outcomes (Thalmayer et al., 2011). The dimensions -- Conscientiousness, Honesty/Propriety, Agreeableness (Kindness & Even Temper), Resiliency vs. Internalizing Negative Emotionality, Extraversion (Gregariousness and Positive Emotionality), and Originality/Talent -- are analogous to Big Five dimensions, with a few adjustments (e.g. Big Six Agreeableness includes aggressiveness and irritability content sometimes found in Big Five Neuroticism). These adjustments allow the Big Six to better map onto core temperament dimensions theorized to underlie mental disorders and adult personality (those proposed by Clark, 2005). The 48-QB6 has been shown to improve on popular Big Five inventories (BFI, NEO-FFI) in terms of predicting important student life outcomes (Thalmayer et al., 2011). Responses are on a 6-point likert scale ranging from “strongly disagree” to “strongly agree”.

Procedure

Undergraduate college students enrolled in introductory psychology and linguistics courses completed surveys using the psychology department’s human subjects pool in exchange for course credit. The 48-QB6 and the OQ-45 were administered to human subjects pool participants as part of a half-hour-long electronic general survey.

Analyses

Confirmatory factor analysis in Mplus 7 using robust maximum likelihood estimation was used to compare the intended structure of the OQ-45 with alternative models identified in Study 1 (three-and four-factor personality models; seven-factor problems models).

Relation between OQ-45 items and scales (including intended structure, personality, and seven-factor subscales) with Big Six dimensions was tested using Pearson correlations. The extent that OQ-45 content is covered by the Big Six model of

personality was tested using multiple regression. Correlation between OQ-45 items and QB6 scales was used to select items for use as personality indicators, as a potential extension of the use of OQ-45 data.

Study 3

The purpose of Study 3 was to explore how well-validated Big Five and Big Six personality constructs might be used to better guide treatment decisions and assess treatment efficacy. Resiliency (vs. Neuroticism) and Extraversion were hypothesized to predict initial level of psychological functioning, in terms of lower total OQ scores. Honesty/Propriety and Conscientiousness scores are also predicted to relate to lower initial OQ scores, to a lesser degree. Personality attribute dimensions are also hypothesized to provide prediction of three therapy usage and outcome variables: number of sessions attended, termination outcome, and change in OQ-45 scores. Specifically, scores on Conscientiousness, Honesty/Propriety, and Agreeableness are hypothesized to predict attending more sessions, scores on Conscientiousness and Honesty/Propriety are hypothesized to predict successful termination, and scores on Openness, Conscientiousness and Honesty/Propriety are hypothesized to predict improvement in psychological functioning (change in OQ-45 scores) over time.

Participants

A total of 255 clients joined the study at the time of their intake at the CFT clinic and completed the "Traits in the Clinic" questionnaire. Of these, 151 had completed therapy and files were available for review as by May 2013. For another 71 clients, therapy is in progress but demographic information and multiple scores on inventories were available and are included in some analyses, as noted below. The full sample (N=222) was 53% female (two clients identified as transgender), and 88% white. Age ranged from 18 to 72, with an average of 34.85 (SD = 11). Four to seven persons (2-3% of the sample) indentified as each of the following: Native American, Asian or Pacific Islander, African American, Hispanic, mixed race, and other. Forty-three percent of the sample identified as Christian, 4% as Jewish, 3 % as Buddhist, 12% as other or spiritual, and 34% as not having a religion. The vast majority of clients identified as heterosexual (82%), 2% as gay or lesbian, 8% as bi-sexual, and 2% as other (7% of the sample declined to answer this question). Almost all clients (95%) reported English as their

native language. The sample is relatively low income, with 91% of clients reporting less than $50,000 a year in household income, and 67% less than $30,000 a year. The sample was relatively well educated, however, with 64% reporting at least some college education, and 13% a graduate degree.

The most common type of therapy obtained for this sample was couples therapy (45%); 32% came to the clinic for individual therapy, and 13% for family therapy. Ten percent of clients engaged in individual therapy in addition to couples or family therapy. The number of sessions attended ranges from one (intake session only) to 58; for those that have completed therapy, the average number attended was 8 (SD = 8).

Materials

The Outcome Questionnaire-45 version 2, as described above. The total score was used in the current study.

The Big Five Inventory (BFI) plus an 18-item addendum (for a total of 62-items administered) to create a 54-item, six-factor version of the BFI (as in Thalmayer et al., 2011). This includes the addition of an Honesty/Propriety factor (highly similar to QB6 Honesty/Propriety), and a Big Six Agreeableness factor, including more content related to (lack of) aggressiveness and irritability than typical Big Five (and BFI) agreeableness.

The three-page "Client Questionnaire" administered to all new clients at the CFT clinic asks clients to provide demographics and background information including age, gender, race, religion, sexual orientation, native languages, level of education, level of employment, and household income, and to answer yes or no to 18 items regarding life stressors and experiences.

Outcome variables in addition to changes in OQ-45 scores include total number of sessions attended and reason for termination.

Procedure

New clients at the CFT clinic who decided to participate in the study for a $5 incentive completed the personality questionnaire and allowed the researcher to access de-identified responses to clinic surveys completed at that time and later if services at the clinic are continued. Time-1 OQ-45 scores, client questionnaire, total number of sessions attended and reason for termination were be available for all participants who complete the survey. Additional OQ-45 scores were available from participants who continued to

receive services at the CFT clinic and completed it at future sessions. Responses on the OQ-45 were associated with date of administration for a time "clock" and with session number for a therapy "clock".

Total number of sessions and reason for termination were added to the dataset from notes in the original client files. Reason for termination was coded into four main categories by the author and a research assistant from the therapist's termination notes, as detailed in Table 3.1. The termination coding system was developed by the author in a large set of existing data obtained from the same clinic (Thalmayer & Baune, 2013). Subcategories were also coded for an expanded alternate set of 14 possible termination outcome descriptions.

Table 3.1. Termination outcome types.

Termination type		Specification	
1	Drop out, *n =54*	10	No-show, disappeared, didn't return calls (1-3 sessions)
		11	Client communicated desire to terminate -- therapist doesn't agree (1-2 sessions)
		12	Couple broke up, discontinued therapy (1-2 sessions)
		13	Refused therapist assigned or made demands clinic couldn't meet
2	Quit before making gains, *n =24*	20	No-show, disappeared, didn't return calls (4+ sessions)
		21	Client communicated desire to terminate -- therapist doesn't agree (3+ sessions)
		22	Decided to go to therapy at a different clinic
		23	Couple broke up, discontinued therapy (3+ sessions)
3	Some goals met, *n =51*	30	No show (after some progress made)
		31	Client moved away
		32	Client can no longer afford services
		33	Therapist left clinic and client decided not to transfer
		34	Client transferred but quit after 1-3 sessions with new therapist
		35	Client feels they have made enough progress.
		36	Scheduling difficulty
		37	Treatment in progress; some goals have been met
4	Treatment complete, *n = 34*	40	Goals met, therapy completed
		41	Borderline case -- better than a 3 but not quite a 4

Analyses

Pearson correlation between OQ-45 total score and BFI-six dimensions was used for initial assessment of the relation between personality attributes and clinical presentation. These analyses came from samples of between 207 to 216 participants, as twenty participants were missing a score on one or more personality scales.

Poisson regression with robust estimators was used to predict sessions attended because this was a highly positively skewed count variable. This analysis included those persons who had completed therapy, and for whom scores on personality variables and all variables included in the final model were available (N= 136).

Ordinal logistic regression was used to predict termination outcome. The sample size varied by model (N=141-163), due to changes in listwise deletion as variables were added to more complex models. For both Poisson regression on number of sessions attended and ordinal logistic regression for termination outcome, analysis preceded in steps, and three models are reported:

A. An initial unconditional model for comparison
B. A model including all potential control variables (demographics, psychological history) was pruned to include only significant predictors
C. A final model including all personality scales and the significant demographic control variables.

Multilevel Modeling using SPSS Mixed was used to test if personality dimensions predict intercept (initial scores) and slopes (change in scores) of OQ-45 total scores in the hypothesized directions.

CHAPTER IV

RESULTS

Study 1

The time-1 dataset included 1,822 responses completed before an intake session. Total scores on the OQ ranged from 13 to 157 (M = 69.8, SD = 23.4). Table 4.1 displays the psychometric properties of the scales used in analyses, in the full time one sample. In addition to Cronbach's alpha and average interitem correlation for each scale, the variance of interitem *r* is included as an indicator of unidimensionality (Ideally, correlations between items measuring a single construct should range from .15 to .50 [Clark & Watson, 1995] -- a larger range in correlations indicates multidimensionality. If variance is zero, a factor analysis will only find one nonsinglet factor.)

The 45 items were moderately correlated with one another in the expected directions (r = .00-.61; average inter-item correlation = .25), with one exception: Item 14 was correlated in the intended direction with only three items, was correlated in the wrong direction with one, and was otherwise uncorrelated with items, a pattern observed in other studies of OQ-45 structure.

Skew and kurtosis values were significant for most items. In the most extreme cases, this was due to low base rate – most participants answered "never" to substance use, suicide, workplace violence, and phobia items, and "almost always" to item 1. Only two substance use items had standard deviations below .6 and absolute skew and kurtosis values likely to be problematic per Kline (2011; SI > 3, KI > 10).

Exploratory Factor Analysis

Approximately one third of responses (N = 624) were randomly selected from the 1,822 intake OQ-45s for the first sample. Exploratory factor analyses used principal axis factoring with oblique (oblimin) rotation and listwise deletion for missing values.

A three-factor structure was extracted to compare the observed structure of the OQ to the intended structure and the three-factor personality model. Table 4.2 reports the correlations between observed factors and hypothesized scales. The content of the first observed factor was interpretable as either Symptom Distress or Negative Affectivity, and correlated highly with both (.98). The second observed factor was slightly less interpretable as Interpersonal Relations than as Positive Affectivity (-.88 vs. -.93). Most

Table 4.1. Descriptive statistics for the OQ-45 intended and hypothesized scales.

Scale (number items)	Mean	SD	α	Mean interitem r	variance interitem r
Total Score (45 items)	69.79	23.37	.94	.247	.019
Intended Subscales*					
Symptom Distress (25)	1.60	.62	.93	.335	.016
Interpersonal Relations (11)	1.64	.53	.80	.260	.024
Social Role (9)	1.34	.58	.66	.203	.023
Three-factor Personality Subscales					
Negative Affectivity (24)	1.76	.62	.92	.327	.010
Positive Affectivity (10)	1.69	.69	.86	.375	.019
Disinhibition (9)	.82	.41	.63	.173	.019
Four-Factor Personality Model Subscales					
Negative Affectivity (24)	1.72	.64	.92	.337	.009
Positive Affectivity (6)	1.58	.74	.84	.463	.006
Affiliation (10)	1.74	.62	.81	.299	.020
Disinhibition (5)	.62	.43	.54	.195	.023
Internalizing/Externalizing Scales					
Internalizing (35)	1.74	.60	.94	.314	.012
Externalizing (8)	.70	.43	.67	.211	.016
OQ-45 Seven-Factor Problems Scales					
Depressive Thinking (9)	1.66	.76	.89	.472	.005
Positive Affectivity (6)	1.56	.75	.84	.478	.008
Anxiety (9)	1.50	.68	.81	.328	.006
Work/School (5)	1.94	.65	.68	.262	.040
Family/Relation Stress (7)	2.08	.74	.79	.353	.014
Conflict (5)	1.08	.55	.71	.324	.007
Substance abuse (3)	.24	.47	.65	.387	.004

Note. N = 1630 - 1810

*Means for the subscales were divided by number of items for an average score on the item level, 0-4 scale, to allow for comparison between subscales.

of the scale's reverse-keyed items loaded onto this factor. The third observed factor included content related to arguments, anger, and substance use, and was thus less interpretable as Social Role than Disinhibition (.78 vs. .82).

Overall the average absolute correlation between observed factor scores and the content-matched intended OQ scales (after an r to z transformation for averaging, then transformation of average back to r) was slightly lower (.92) than for personality scales (.94), but overlapping 95% confidence intervals for the averages indicate a non significant difference. The average fit of four-factor solution factors to respective four-

Table 4.2. Correlations between EFA factors and intended and hypothesized OQ-45 scales.

Factor	Subscales						
			Two factor solution				
	INT	EXT					
F1 of 2	**93***	66*					
F2 of 2	-74*	**-23***					
			Three factor solution				
	SD	IR	SR	NA_3	PA_3	D_3	
F1 of 3	**98***	61*	67*	**98***	67*	47*	
F2 of 3	-68*	**-88***	-39*	-66*	**-93***	-25*	
F3 of 3	54*	43*	**78***	51*	51*	**82***	
			Four factor solution				
	NA_4	AF_4	D_4	PA_4			
F1 of 4	**99***	57*	34*	67*			
F2 of 4	-63*	**-88***	-10*	-79*			
F3 of 4	30*	32*	**86***	11			
F4 of 4	-46*	-25*	-38*	**-66***			
			Seven factor solution				
	DT	FRS	SA	PA	WSA	Con	Anx
F1 of 7	**94***	51*	19*	64*	52*	53*	85*
F2 of 7	-54*	**-95***	-12*	-66*	-34*	-25*	-37*
F3 of 7	35*	28*	**95***	31*	24*	38*	30*
F4 of 7	-62*	-32*	-15*	**-82***	-39*	-74*	-36*
F5 of 7	52*	-20*	.09	45*	**88***	45*	49*
F6 of 7	-13*	04	11*	-25*	17*	**57***	11
F7 of 7	56*	47*	11	43*	54*	45*	**76***

Note. The first random third of the time one sample was used, with listwise deletion for those missing scores on OQ-45 items (N=530). Decimal points removed for readability. Expected matches are bolded – the averages reported in text are of the bolded values. INT = internalizing, EXT = externalizing; SD = Symptom Distress, IR = Interpersonal Relations, SR = Social Role. NA = Negative Affectivity, PA = Positive Affectivity, D = Disinhibition; AF = Affiliation. DT = Depressive Thinking, FRS = Family/Relationship Stress, SA =Substance Abuse, WSA = Work/School Adjustment, Con = Conflict, Anx = Anxiety.

* Correlation is significant at $p < .01$.

factor personality scales was similar (.91). The main source of misfit was a mixing of Affiliation and Positive Affectivity items across the third and fourth factors.

The average fit of two-factor solution factors to internalizing and externalizing scales was weaker (.74). While the first factor was interpretable as internalizing tendencies, the second did not emphasize externalizing tendencies. None of the eight

items chosen *a priori* as belonging to the domain loaded most highly on the observed second factor.

The initial seven-factor model subscales were difficult to match one-to-one to the seven-factor EFA solution in terms of primary content. Modifications to this model were made, drawing on multiple logical alternatives of item placement and subscale structure identified *a priori* in the sorting process. (See Appendix for more details on the developments of the alternative subscales.) A process of extracting additional factors in a series until they became uninterpretable was used to fully explore the structure of the data, and how it matched potential subscale structures. A ten-factor solution was interpretable as a fully elaborated model of specific problems. A large factor of 'Depressive Thinking' items (about 29.5% of the variance) was followed by factors interpretable as 'Relationship Malaise', 'Substance Abuse', 'Work/School Adjustment', 'Family Trouble', 'Somatic Depression', 'Anxiety', 'Positive Emotionality', 'Conflict', and item 14, which loaded only its own factor. Many changes implied by this empirical solution matched logical alternatives identified *a priori*. Thus, the empirical model was revised, named as above, with a few exceptions: To avoid having scales with fewer than 3 items, relationship issues and family troubles were combined onto a single 'Family Relationship Stress' scale; somatic depression items had very high secondary correlations with Depressive Thinking and with Anxiety, and thus were matched with their secondary scales; and item 14 was grouped with 'Work/School Adjustment' for rational reasons, despite its lack of correlation with them. Average correlation between observed factors and analogous revised scales was .88.

Confirmatory Factor Analysis (CFA)

A baseline total score model, the intended subscale model and the two personality and specific problem structural models were compared in the second random third of the intake dataset (N= 569). All models were run first with factors allowed to correlate, secondly as bi-level models including a total score factor (as in Coco et al., 2008, and Bludworth et al., 2010). Results of these analyses appear in Table 4.3.

All chi-square values were significant, as expected given the sample size. All models had SRMR values under .08. No model had CFI greater than .90, although the bi-level seven-factor model was close, at .87. Most bi-level models and the oblique seven-

Table 4.3. Summary of CFA fit indices for alternative models of OQ-45 structure.

Model	df	adj. χ^2	CFI	AIC	BIC	TLI_1	RMSEA (p close)	SRMR	$R^2 \geq .5$/ $R^2 < .10$
Second Random Third (N = 569)									
Baseline: Total Score	945	3,790	.711	70,177	70,776	-	.069 *	.068	4/6
Intended structure (oblique)	942	3,426	.748	69,772	70,384	.124	.065	.069	5/4
Intended structure bi-level	897	2,569	.830	68,872	69,684	.382	.055	.053	11/4
Three-factor Personality (oblique)	942	3,151	.776	69,457	70,070	.219	.061	.062	8/4
Three-factor Personality bi-level	897	2,565	.831	68,847	69,659	.392	.055	.051	12/4
Four-factor Personality (oblique)	939	3,100	.781	69,392	70,018	.236	.061	.064	8/2
Four-factor Personality bi-level	894	2,548	.832	68,813	69,639	.379	.054	.052	9/1
Int./Externalizing (oblique)	944	3,566	.734	69,912	70,516	.076	.067	.066	5/5
Int./Externalizing Bi-level	899	2,742	.813	69,048	69,851	.312	.057	.052	10/3
Seven-factor Problems (oblique)	924	2,559	.834	68,814	69,506	.412	.053 (.02)	.059	11/1
Seven-factor Problems Bi-level	880	2,148	.871	68,398	69,285	.521	.048 (.89)	.046	14/1
Third Random Third (N = 629)									
Baseline: Total Score	945	3,938	.697	72,074	72,674	-	.071	.071	2/8
Intended Structure bi-level	897	2,618	.818	70,621	71,434	.395	.056	.052	10/3
Three-factor Personality bi-level	897	2,692	.809	70,637	71,450	.369	.058	.056	10/3
Four-factor Personality bi-level	894	2,480	.840	70,446	71,273	.440	.053 (.01)	.049	12/1
Seven-factor Problems bi-level	880	2,129	.874	70,065	70,319	.552	.048 (.95)	.046	15/1
Time Five (N = 681)									
Baseline: Total Score	945	4,248	.728	70,813	71,424	-	.072	.067	8/4
Intended Structure bi-level	897	2,855	.839	69,289	70,117	.376	.057	.053	13/3
Three-factor Personality bi-level	897	2,930	.833	69,355	70,183	.352	.058	.052	15/4
Four-factor Personality bi-level	894	2,867	.838	69,251	70,092	.370	.057	.050	13/1
Seven-factor Problems bi-level	880	2,269	.886	68,640	69,544	.549	.048 (.89)	.044	18/1

Table 4.3. (continued).

Model	df	adj. χ^2	CFI	AIC	BIC	TLI_1	RMSEA (p close)	SRM R	$R^2 \geq .5$/ $R^2 < .10$
Time 10 (N = 392)									
Baseline: Total Score	945	3,001	.709	40,736	41,272	-	.074	.073	6/4
Intended Structure bi-level	897	2,118	.827	39,788	40,514	.376	.059	.059	15/4
Three-factor Personality bi-level	897	2,082	.832	39,744	40,470	.394	.058	.057	14/1
Four-factor Personality bi-level	894	2,016	.841	39,660	40,398	.424	.057	.054	13/1
Seven-factor Problems bi-level	880	1,699	.884	39,333	40,128	.573	.049 (.72)	.049	18/1
Student Sample (N = 589)									
Baseline: Total Score	945	4,071	.636	64,302	64,893	-	.075	.080	3/7
Intended Structure bi-level	897	3,241	.727	63,090	63,891	.211	.067	.065	9/5
Three-factor Personality bi-level	897	2,686	.792	62,705	63,506	.398	.058	.061	9/3
Four-factor Personality bi-level	894	2,555	.807	62,555	63,370	.439	.056	.059	11/4
Seven-factor Problems bi-level	880	2,372	.826	62,301	63,177	.488	.054 (.01)	.055	12/2

Note. Analyses run in Mplus 7 using MLR estimation. All adjusted χ^2 values $p < .01$. CFI = comparative fit index; TLI_1 = Tucker-Lewis index comparing each model to total score model; RMSEA = root mean square error of approximation; SRMSR = standardized root mean square residual. Last column is a count of items with R^2 value indicating a majority of variance accounted for by model, contrasted with a count of variables with variance not well accounted for by model.

* P close values were < .001 unless p value is listed.

factor model had RMSEA values under .06 -- in the case of the seven-factor models, the value did not differ from .05, indicating close fit.

More factors increased fit across alternatives, and bi-level models fit better than oblique models. This is illustrated in values of TLI_1 (improvement in fit over total score), which ranged from .08/.31 for oblique/bi-level internalizing-externalizing model, and .12/.38 for the intended structure to .22/.39 and .24/.38 for personality models and .41/.52 for the seven-factor model. AIC and BIC values, which adjust for parsimony, also indicate that bi-level four- and seven-factor models were the best fitting.

In the only case where models with the same number of factors were compared, the three-factor personality model appeared superior to the intended structure (oblique/bi-level values: TLI_1 .22/.39 vs. .12/.38; eight/twelve vs. five/eleven items with variance well explained; AIC 69,457/68,847 versus 69,772/68,872).

In comparing the two personality models, EFA results indicated a slight advantage for the three-factor version. The evidence in CFA is mixed – the four factor versions have slightly lower AIC (69,392/68,813 vs. 69,457/68,847), but BIC values hardly differ, other indices are mixed, and more items are well explained by the three-factor bi-level model than its four-factor counterpart. Thus, both personality models were retained for subsequent analyses.

The seven-factor model of problems had a clear advantage over the internalizing externalizing model in both oblique and bi-level versions. Thus, the latter was dropped from subsequent analyses.

Preferred Models CFA

The total score baseline model and bi-level intended structure, personality, and seven-factor problems models were compared in the last random third of the intake dataset (N = 629). Fit indices are reported in Table 4.3. The pattern of results is highly similar to that observed in the prior sample: The seven-factor model approaches good fit by standard benchmarks, followed by the four-factor personality model. (In this set both the intended structure and four-factor models have very slight advantages over the three-factor personality model.)

An exploration of standardized parameter estimates in the four- and seven-factor models indicates that in both cases, in oblique models, items loaded highly and

consistently on their intended factors. In the bi-level models, however, one latent factor in each model (Negative Affectivity and Anxiety, respectively) had much lower (and some negative) loadings after the addition of the total score factor.

Model Comparison in Time-5 and Time-10 Datasets

The time-5 dataset included 614 OQ-45 responses completed prior to a client's fifth or sixth session. Total scores ranged from 7 to 143 (M = 62.11, SD = 23.12). Scores on most items appeared to be reasonably normally distributed, with the exception of items 11 and 32, as noted above per the time one dataset. The time-10 dataset included 392 OQ-45 responses completed prior to a client's 10th, 11th, or 12th therapy session. Total scores range from 7 to 130 (M = 62.57, SD = 23.36). The distribution of scores was again reasonably normal, excepting items 11 and 32.

The seven-factor model was also the best fitting in the time-5 and -10 datasets. In the time-5 set, there was no clear advantage for any other model. In the time-10 set, There was a slight advantage for the four-factor personality model over the three-factor version, which had a slight advantage over the intended model.

Study 2

Table 4.4 presents descriptive statistics for the inventories administered to the student sample. In the student sample, all 48-QB6 items and scales appeared to be normally distributed. While many OQ items in this sample appeared somewhat skewed or kurtotic on visual inspection, no items had absolute skew and kurtosis values likely to be problematic per Kline (2011; SI > 3, KI > 10). The lowest item standard deviations were .74 (item 11) and .77 (39). Total scores ranged from 12 to 144 (M = 60.73, SD = 21.14). The mean for this college sample was noticeably higher than the manual's undergraduate norms of 42 to 51 (drawn from classroom samples in Utah, Idaho, and Ohio; Lambert, Morton, et al., 2004), and was more similar to the mean of the older, clinical sample in Study 1 (62.75). The clinical sample scored more problematically than the student sample on 31 items. But perhaps in part due to the age difference between the clinical and student samples, the student sample scored more problematically on 12 items (1, 4, 11, 12, 14, 22, 26, 32, 35, 38, 39, 44), including substance use, work violence, phobia, and all work/school stress items.

Table 4.4. Descriptive statistics for scales in college student sample.

Scale	Mean	SD	α	Mean interitem r	variance interitem r
		48QB6			
Conscientiousness	3.15	.79	.74	.273	.009
Honesty/Propriety	3.12	.81	.73	.259	.011
Agreeableness	2.84	.73	.70	.219	.008
Resiliency	2.85	.78	.72	.237	.020
Extraversion	3.51	.74	.74	.270	.017
Originality/Talent	3.30	.66	.65	.192	.008
		OQ-45			
Total Score (45 items)*	60.73	21.14	.93	.240	.020
		OQ-45 Intended Subscales			
Symptom Distress (25)	33.38	12.93	.91	.293	.016
IR (11)	14.67	6.29	.76	.241	.023
Social Role (9)	12.42	4.04	.64	.173	.028
		OQ-45 Four-Factor Personality Model subscales**			
Negative Affectivity (24)	1.50	.52	.90	.28	.014
Positive Affectivity (6)	2.80	.64	.83	.46	.015
Affiliation (11)	2.76	.55	.76	.25	.023
Disinhibition (5)	.65	.62	.77	.42	.012
		OQ-45 Seven-Factor Psychopathology scales			
Depressive Thinking (9)	1.28	.64	.87	.424	.010
Positive Affectivity (6)	1.19	.66	.85	.481	.014
Anxiety (9)	1.34	.59	.80	.306	.008
Work/School (5)	2.02	.53	.58	.188	.033
Family/Relation Stress (7)	1.45	.66	.74	.291	.024
Conflict (5)	1.04	.54	.51	.231	.038
Substance abuse (3)	.59	.72	.73	.509	.003

Note. (N = 545-577 for QB6 scales, N=511-588 for OQ scales).

*Number of items is noted in parentheses for OQ scales.

** Means for alternative subscales were divided by number of items for an average score on the item-level, 0-4 scale to allow for better comparison between subscales.

Results of the tests of fit for the preferred models of OQ-45 structure are reported at the bottom of Table 4.3. Again, indices suggest best fit for the four-and seven-factor models. Table 4.5 provides correlations between the scales. It was hypothesized that Resiliency, Extraversion, Honesty/Propriety and Conscientiousness, in that order, would correlate most highly with the OQ total score. Indeed, Resiliency had the highest

magnitude correlation with total score (-.59), followed by Extraversion (-.41). However, correlations between total score and Honesty/Propriety and Conscientiousness were similar in magnitude, and that with Agreeableness was close behind (-.31-.34). Not surprisingly, Resiliency correlated most strongly with Symptom Distress (-.63), Negative Affectivity (-.62), and Depressive Thinking (-.59) subscales.

Table 4.5. Correlations between QB6 scales and OQ-45 scales.

	QB6 scales					
	R	E	H	C	A	O
OQ Total	**-59**	**-41**	-33	-34	-31	-24
Symptom Distress	**-63**	-37	-22	-25	-24	-21
Interpersonal Relations	**-40**	-37	-33	-37	-26	-19
Social Role	**-43**	-29	-30	-29	-38	-30
Three-factor Personality						
Negative Affectivity	**-62**	-30	-18	-22	-21	-14
Positive Affectivity	**49**	**45**	29	37	23	27
Disinhibition	-26	-32	**-41**	-25	-38	-29
Four-factor Personality						
Negative Affectivity	**-62**	-30	-19	-23	-24	-16
Positive Affectivity	**53**	**46**	23	32	24	34
Disinhibition	*-05*	*-10*	**-45**	-28	-18	-14
Affiliation	**40**	**41**	35	37	31	25
Problem scales						
Depressive Thinking	**-59**	-36	-18	-27	-19	-20
Positive Affectivity	**52**	**46**	24	33	22	31
Anxiety	**-51**	-26	-15	*-06*	-18	-13
Family/Relation Stress	**-41**	-28	-23	-28	-25	-15
Substance Abuse	*-05*	*-10*	**-45**	-28	-18	-14
Work/School Adjustment	**-43**	*-11*	-13	-31	-25	*-11*
Conflict	-39	-38	-27	-26	**-40**	-27

Note. N = 484 -507. Decimal points removed. Correlations not significant at $p < .01$ (2-tailed) are italicized. Correlations .40 or greater are in bold. R = Resiliency, E = Extraversion, H = Honesty, C = Conscientiousness, A = Agreeableness, O = Originality.

A multiple regression indicated that the six QB6 scales explain about half the variance in OQ total score ($F [6, 443] = 64.31$, $p < .001$, $R^2 = .46$).

In order to determine if OQ items might be used as markers for personality dimensions, correlations between OQ items and QB6 scales were explored. Correlations

of .10 in magnitude and higher (significant at $p < .01$) are shown in Table 4.6, and items selected as personality markers are bolded. The scales defined by OQ items are narrower than full personality attribute scales (with the possible exception of the NA), and are better defined as subcomponents of the broader constructs: Negative Affectivity (17 items), Positive Affectivity (6 items), Substance Abuse (3 items), Work/School Adjustment (3 items), and Hostility (5 items). Some distinctions between PA and NA items were difficult to make – a factor analysis of candidate items for the two scales, conducted in the full intake clinical sample, led to final choices on items that loaded highly on both scales here. Table 4.7 shows correlations between the OQ-45 personality subscales and the full personality dimensions in the 48QB6.

Table 4.6. OQ-45 items for personality scales and correlation with QB6 Dimensions.

OQ-45 Item	R	E	H	C	A	O
36. I feel nervous.	**-.51**	-.22				
15. I feel worthless.	**-.50**	-.30		-.24	-.16	-.17
10.	**-.49**	-.21		-.12	-.12	-.14
9.	**-.44**	-.27	-.12	-.17	-.12	-.14
33.	**-.44**	-.19	-.16	-.14	-.18	
42.	**-.44**	-.22		-.20		
23.	**-.43**	-.24	-.19	-.27	-.20	-.22
18.	**-.41**	-.21		-.17		
5.	**-.40**	-.11				
4.	**-.39**		.12		-.18	
40.	**-.38**	-.28	-.22	-.18	-.12	
2.	**-.37**	-.11		-.12	-.11	
20.	**.35**	.30	.21	.25	.14	.17
25.	**-.33**	-.12	-.17		-.17	
29.	**-.33**	-.17	-.11		-.16	-.16
45.	**-.33**	-.15			-.15	
21.	.31	.27	.11	.15	.14	.26
27.	-.28	-.18	-.12			-.11
41.	-.25		-.15			
16.	-.21					
7.	-.20	-.15	-.16	-.18	-.20	
43. I am satisfied with my relationships with others.	.37	**.39**	.20	.27		.19
31. I am satisfied with my life.	.48	**.38**	.17	.21	.20	.24
24.	.47	**.37**	.15	.22	.11	.26
13.	.46	**.47**	.21	.25	.23	.24
3.	-.35	**-.35**	-.11	-.20	-.15	-.27
35.	-.29	-.31				-.20

Table 4.6. (continued).

OQ-45 Item	R	E	**H**	C	A	O
8.	-.26	-.29	**-.25**	-.15	-.15	-.17
1.	.11	.22		.16		
26. I feel annoyed by people who criticize my drinking (or drug use). *			**-.40**	-.20	-.14	
32. I have trouble at work/school because of drinking or drug use. *		-.12	**-.38**	-.26	-.14	-.13
11.		-.15	**-.30**	-.24	-.16	-.19
22. I have difficulty concentrating.	-.21		-.19	**-.40**		
12. I find my work/school satisfying.	.26	.20	.25	**.39**	.23	.23
38.	-.24	-.14	-.21	**-.32**	-.25	-.15
14.	-.12		.21	.26		
37.	.21	.17	.15	.26		
17.	-.14	-.13	-.12	-.22		
28.	-.21		-.13	-.21	-.16	
19. I have frequent arguments.	-.25	-.19	-.27	-.13	**-.45**	-.26
6. I feel irritated.	-.39	-.13			**-.38**	-.15
39.	-.23	-.26	-.34	-.22	**-.31**	-.25
30.	-.34	-.37	-.24	-.20	**-.31**	-.27
44.	-.17	-.25	-.33	-.16	**-.30**	-.27

Note. Items are sorted by personality scale (items used in scales are bolded), then in order of factor loadings. The two highest loading items are shown to give an indication of scale content. Correlations ≤.10 are blanked. Items are sorted in order of factor loadings. Items selected for personality scales are bolded under relevant column. R = Resiliency, E = Extraversion, H = Honesty, C = Conscientiousness, A = Agreeableness, O = Originality.

* Item ends "If not applicable, mark 'Never'."

Table 4.7. Correlations between personality subscales derived from OQ-45 items and QB6 scales.

	QB6 scales					
	R	E	H	C	A	O
Neuroticism	**-.66**	-.29	-.12	-.17	-.24	-.14
Positive Affectivity	.55	**.51**	.30	.30	.24	.31
Substance Abuse	*-.05*	*-.10*	**-.45**	-.28	-.18	-.14
Work School Adjust.	.31	.14	.28	**.49**	.26	.19
Hostility	-.39	-.34	-.38	-.22	**-.50**	-.34

Note. All correlations significant at $p < .01$ except those italicized. Intended correlations are in bold. R = Resiliency, E = Extraversion, H = Honesty, C = Conscientiousness, A = Agreeableness, O = Originality.

Study 3

Clinical Presentation

The correlations between OQ-45 total score and personality dimensions are displayed in Table 4.8, along with scale statistics for the personality scales. Neuroticism (positively) and Extraversion (negatively) were hypothesized to predict initial OQ scores. Honesty/Propriety and Conscientiousness were predicted to relate to lower initial OQ scores to a lesser degree. While Neuroticism had the highest correlation with the OQ-45, and Openness the lowest, Extraversion did not relate more strongly to the OQ than did the other scales. There is overlap between confidence intervals for the correlations of the OQ with Agreeableness, Conscientiousness, Extraversion, and Honesty/Propriety.

Table 4.8. Descriptive statistics for BFI-Six scales and correlation with intake OQ-45.

Scale	M	SD	α	Mean interitem r	Variance interitem r	*r* with OQ-45	Lower 95% CI	Upper 95% CI
Neuroticism	3.34	.79	.84	.39	.008	.69**	.62	.76
Agreeableness-six	3.32	.73	.78	.31	.012	-.41**	-.52	-.29
Agreeableness	3.71	.60	.75	.26	.008	-.35**	-.47	-.23
Conscientiousness	3.62	.66	.79	.31	.007	-.39**	-.50	-.27
Honesty/Propriety	3.63	.60	.68	.18	.011	-.24**	-.37	-.11
Extraversion	3.30	.85	.85	.41	.017	-.22**	-.35	-.09
Openness	3.81	.65	.79	.29	.021	-.02	-.16	-.11

N = 207- 222. BFI items are measured on 1-5 scale, and averaged here to be on 1-5 scale.

** $p < .01$

Table 4.9 reports correlations between BFI-six scales. As intended the Big Six Agreeableness factor is less highly intercorrelated with Conscientiousness and Honesty/Propriety. This factor is included in subsequent analyses, in place of regular BFI Agreeableness.

Overall scores were hypothesized to show the same pattern of correlation in this clinical sample as in the non-clinical sample in Study 2, although a different measure of the Big Six personality model was used (48 QB6 versus the BFI-six). While the rank order of correlations for scales in the student sample (QB6) differed slightly (Resiliency,

Table 4.9. Correlation matrix for personality scales (N = 201-212).

	N	E	C	H	A	A6
Extraversion	-.25**					
Conscientious	-.35**	.10				
Honesty	-.23**	-.17*	.41**			
Agreeable	-.39**	.09	.30**	.32**		
A6	-.49**	-.01	.19**	.24**	.70**	
Openness	-.01	.22**	-.09	-.07	.05	.13

* p < .05; ** p < .01

Extraversion, Conscientiousness, Honesty/Propriety, Agreeableness, Originality) from those in the clinical sample, the differences were not statistically significant. In both cases the dimension related to negative affectivity (Resiliency and Neuroticism, respectively) had the highest correlations with the OQ-45 total score, and Openness/Originality the lowest; the correlations between the OQ-45 and remaining scales had overlapping 95% confidence intervals (after r to z transformations). (In the student sample, the correlation with Originality did not differ significantly from those with Agreeableness, Conscientiousness, or Honesty/Propriety.)

Sessions Attended

Number of sessions attended was explored in participants who had completed therapy at the time of analyses. Average sessions for this sample are reported in Table 4.10. Categorical control variables were initially explored using ANOVA and post hoc contrasts, to determine useful dummy codes for subsequent analyses.

Type of therapy was associated with number of sessions attended, $F(3, 147) = 8.81$, $p < .01$. Scheffé post hoc tests indicated that clients who received multiple types of therapy attended significantly more sessions (M=16.81, SD= 14.14) than those who attended only one type of therapy (see Table 4.9; $p < .01$ across three comparisons). Thus, a dummy code was used to account for therapy type (mixed types versus one type) in analysis of sessions attended. Sessions appeared to vary by marital status ($F [4, 140] = 2.97$, $p = .02$), although no post hoc comparisons were significant. A contrast capturing the largest difference, between those who reported being married (M = 10.12, SD = 10.32) and everyone else (M = 5.62) was included in the analysis. Sessions did not vary by ethnicity ($F[5, 139] = .25$, $p = .94$), gender ($F [2, 75.06] = .56$, $p = .57$), or sexual

orientation (F[3, 135] = .67, p = .57), and no related post hoc comparisons were significant. These variables were excluded from subsequent analyses.

Table 4.10. Mean Sessions by Therapy Type.

Therapy Type	N	Mean	Std. Deviation
Individual	55	5.95	5.48
Couples	57	7.40	7.77
Family	23	6.83	4.54
Mixed	16	16.81	14.14
Total	151	7.78	8.16

Table 4.11 reports results for three models of predictors of sessions attended using Poisson regression: an unconditional model including only the intercept, a model including only the demographic variables found to be significant predictors of sessions attended, and a final model including significant control variables and personality scales. Demographic descriptors that were excluded after an initial run because they did not appear to influence number of sessions attended after accounting for type of therapy, age, marital status, and level of education, included income and gender.

For an unmarried 18 year old with an 11th grade education who attended the clinic for individual or couples therapy, the average number of sessions attended was 2.1. For those who attended more than one type of therapy (individual plus either couples or family sessions), this average was more than doubled (rate = 2.4), and it increases nearly that much (rate = 1.6) for those who are married. The average also increases slightly for each year older a client is at intake (rate = 1.02) and for each additional year of education(rate = 1.1). Accounting only for these control variables in Model B explained 29% of the total variance in number of sessions attended. Scores on Conscientiousness, Honesty/Propriety, and Agreeableness were hypothesized to predict attending more sessions, but no personality scales significantly predicted sessions attended.

To compare the validity of the OQ-45 to that of the Big Six in terms of predicting clinical outcomes, another set of analyses used OQ-45 scores in place of personality scales for an alternative final model. Total score was a significant predictor of sessions (B = .01, rate = 1.01, p = .024), and the model explained a similar amount of variance to the

Table 4.11. Predictors of sessions attended.

	Unconditional model			Control variables				Personality model			
Predictor	B	SE	Wald χ^2	B	rate	SE	Wald χ^2	B	rate	SE	Wald χ^2
Intercept	2.10	.09	539.61***	1.09		.20	30.55***	2.08		1.32	2.50
Multiple types therapy				.81	2.25	.18	19.75***	.88	2.41	.18	24.08***
Age (years over 18)				.02	1.02	.01	4.67*	.02	1.02	.01	8.69**
Married				.52	1.68	.14	14.42***	.47	1.60	.15	9.59**
Education (years over 11)				.09	1.09	.04	5.87*	.12	1.13	.04	9.28**
<u>Personality Variables</u>											
Conscientiousness								.01	1.01	.13	.01
Honesty/Propriety								-.14	.87	.19	.52
Agreeableness (BFI six)								-.05	.95	.12	.16
Extraversion								-.06	.94	.09	.43
Neuroticism								-.07	.93	.13	.26
Openness								-.14	.87	.13	1.11
<u>Goodness of fit</u>											
R^2_L				.29				.32			
Deviance (df)	932.06 (132)			657.33 (128)				630.39 (122)			
AIC	1401.01			1134.28				1119.34			
BIC	1403.90			1148.73				1151.13			

N = 135 terminated clients with complete data on all variables included in final model. Rate = exp(B), the rate of increase for sessions for each unit change in predictor. $R^2_L = R^2$ for logistic regression models per Cohen et al. (2003).
~ p < .10; * p < .05; ** p < .01; *** p < .001.

model including personality scales (deviance = 619.51; unconditional model with appropriate listwise deletion, N = 141, deviance = 940.48, R^2 = .34). When the three OQ-45 subscales were used in place of total score, the interpersonal relations scale significantly predicted sessions (B = .43, rate = 1.53, p < .001) and 42% of variance in sessions was accounted for (deviance = 526.17; unconditional model with adjusted listwise deletion, N = 135, deviance = 909.15).

Therapy Outcome

For clients who had terminated by the time of analyses, 23% had successfully met their goals, and 26% left after experiencing at least some progress. Of the remaining 52% who did not appear to make progress in therapy, the majority (36%) dropped out after one to three sessions. For clients still in therapy at the time of analyses, 17% (12 clients) were rated as clearly making progress (no conclusive information was available for the remainder). Rates of success did not differ significantly by type of therapy, $\chi^2(9) = 10.10$ p = .34. Marital status ($\chi^2[15] = 13.73$, p = .55), sexual orientation ($\chi^2[9] = 12.82$, p = .17), ethnicity ($\chi^2[15] = 16.16$, p = .37), or religious affiliation ($\chi^2[12] = 18.65$, p = .10). Table 4.12 reports results for three models of predictors of therapy outcome using ordinal logistic regression: an unconditional model including only the intercept, a model including significant demographic variables, and a final model including significant control variables and all personality scales. Demographics excluded from Model B due to non-significance included age and type of therapy. Sample size varied slightly between the models because information on one or more demographic variables was missing for ten participants, and one or more personality scores were missing for 12. Scores test of proportional odds indicate that the slope coefficients did not significantly differ across response categories in any of the models.

Female clients and those with more education and income were observed to achieve slightly better outcomes. Together, these variables accounted for 8% of the total variance in categorical therapy outcome. Scores on Conscientiousness and Honesty/Propriety were hypothesized to predict successful termination, but after accounting for gender, education, and income, no personality variables predicted number of sessions attended. Together, the six personality scales appeared to account for about 2% of the total variance in therapy outcome.

Table 4.12. Predictors of termination outcome.

	Unconditional model			Demographics only				Final model			
Predictor	B	SE	Wald	B	SE	Wald	OR	B	SE	Wald χ^2	OR
Threshhold = 1	-.70	.17	17.81***	.06	.40	.02		-2.83	2.40	1.39	
= 2	-.09	.16	.30	.73	.40	3.35~		-2.13	2.40	.79	
= 3	1.33	.19	47.85***	2.29	.44	26.95***		-.56	2.39	.06	
Demographics											
Gender = Male				-.53	.30	3.04~	.59	-.68	.35	3.77~	.51
Education (over 11 yrs)				.18	.08	5.09*	1.20	.21	.09	5.18*	1.23
Income				.11	.05	4.05*	1.12	.11	.06	3.67~	1.12
Personality											
Conscientiousness								-.11	.06	.18	.90
Honesty/Propriety								-.09	.32	.09	.91
Agreeableness (BFI six)								-.13	.25	.24	.88
Extraversion								-.23	.21	1.16	.79
Neuroticism								-.06	.26	.05	.94
Openness								-.20	.28	.49	.82
Goodness of fit											
Pseudo R^2 (Nagelkerke)	-			.08				.10			
-2LL	-			273.64				365.30			
Model χ^2 (df)	-			12.51 (3)**				13.74 (9) *ns*			
Scores test of proportional odds χ^2 (df)				3.01 (6) *ns*				12.61 (18) *ns*			
N	163			153				141			

Note. OR = odds ratio.
~ p < .10; * p < .05; ** p < .01; *** p < .001.

To test the comparative validity of the OQ-45 to the Big Six, another set of analyses used OQ intake scores in place of personality scales for an alternative final model. Total score was a slight and marginally significant predictor of outcome (B = .01, OR = 1.01, p = .08), and the model explained no less overall variance in outcome than the personality model, with fewer predictors (Nagelkerke pseudo R^2 = .11). Using the three subscales instead of total score resulted in no significant predictors, but the same variance explained.

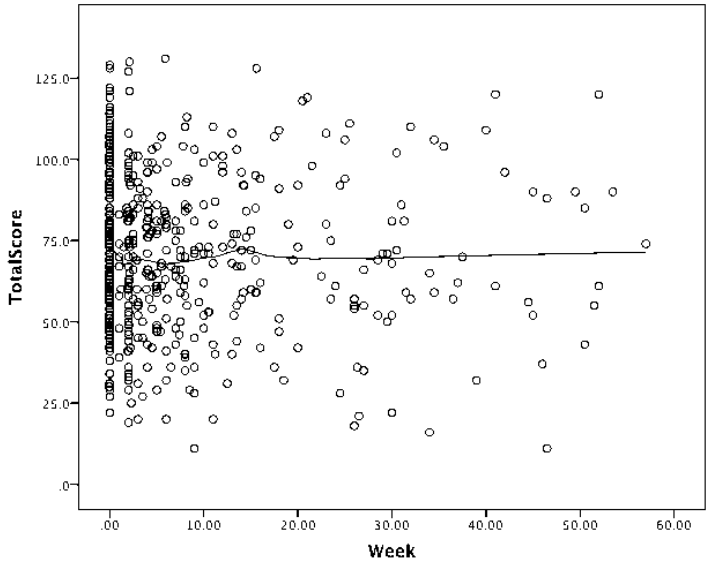
Change in OQ-45 Scores

Clients who attended the clinic for individual, couple, or mixed therapy, and who provided information on the six personality scales (N= 176; this criteria excluded 34 participants) were included in longitudinal analyses on
predictors of change in OQ-45 score. Those who attended solely for family therapy were excluded because for many families, a child (who typically does not complete the OQ-45) provided the impetus for seeking treatment, and treatment focused on parenting strategies. There was thus some ambiguity as to whether to hypothesize broad change. Clients included in analyses attended from one to 58 sessions (M = 11.34, SD = 10.75). The distribution of sessions attended was positively skewed (skew = 1.53, SE = .18, kurtosis = 2.1, SE = .36) -- ninety percent of the sample attended 28 sessions or fewer (8.5% clients attended only one session.)

An initial analysis assessed how to model time. A series of polynomials was created to test for the significant of linear, quadratic, and cubic trends. While the linear trend was significant, models including the quadratic and cubic terms did not converge, likely due to multicollinearity and the small sample size. In a larger dataset from the same clinic, quadratic and cubic terms were observed to not account for significant change in slope (Thalmayer & Baune, 2013). There is every reason to expect a similar trend in the current dataset. Additionally, Figure 4.1 indicates the overall linear trend in the current sample.

Table 4.13 reports results for five models of longitudinal change in OQ-45 scores: (a) an unconditional means model including only the intercept, (b) an unconditional linear growth model, (c) a model including only the time-varying predictor, number of therapy sessions attended since the previous measurement, (d) a model including the

Figure 4.1. Scatterplot of OQ-45 score by week, including Loess fit line.



personality variables that were hypothesized to predict intercept and slope of the OQ-45, and (e) a final model including demographic variables found to be significant predictors of outcome, and all hypothesized and non-hypothesized but statistically significant personality scale predictors of intercept and slope.

The unconditional means model indicates that the average client's OQ-45 scores change over time, and that clients differed from one another in OQ-45 scores (variance components both significant at $p < .001$). Almost three quarters of the differences in OQ-45 scores are between people (ICC = .72; this indicates that about 28% of the variation is across time points); this correlation in the unconditional means model also estimates the autocorrelation (Singer & Willett, 2003), and here indicates a high correlation between pairs of composite residuals for each individual.

The unconditional linear growth model indicates that in general, OQ-45 scores decreased by a small amount per week for these clients (-.26, $p < .01$). Significant variation between clients in initial OQ scores (388.83, $p < .001$), and significant within person variation over time (132.46, $p < .001$) remain, but there does not appear to be significant variation in their change over time (.06, *ns*), and only about 10% of the within-person variation in OQ-45 scores can be accounted for by linear time. Model C included only the addition of a time varying predictor , to account the role of the rate of

Table 4.13. Multilevel models of change in OQ-45 scores.

		Parameter	Model A	Model B	Model C	Model D	Model E
Fixed Effects	Intercept	γ_{00}	69.72	71.04***	71.10***	20.46*	18.02*
Initial status, π_{0i}	EXTRAVERSION	γ_{02}				-1.52	-1.65
	NEUROTICISM	γ_{03}				16.28***	17.14***
Slope, π_{2i}	WEEK	γ_{10}		-.26**	-.22*	.86	1.24
	SESSIONS	γ_{30}			-.00		
	CONSCIENTIOUSNESS	γ_{11}				-.03	-.05
	HONESTY/PROPRIETY	γ_{12}				-.22~	-.27*
	OPENNESS	γ_{13}				-.05	-.01
	NEUROTICISM	γ_{14}					-.18*
	AGE						.01~
Variance Components							
Level 1	Within-person	ε_{ij}	146.82***	132.46***	133.59***	132.50***	130.80***
Level 2	In initial status	τ_{00}	385.10***	388.83***	393.14***	213.42***	204.29***
	In rate of change	τ_{11}		.06	.05	.05	.05
Pseudo R^2 Statistics and Goodness of fit							
	$R^2\varepsilon$			.10	-.01	.00	.01
	R^2_0					.45	.47
	R^2_1				.17	.17	.17
	ICC		.72	.75	.75	.62	.61
	Deviance		4351.49	4332.67	4758.44	4100.42	4080.68
	AIC		4357.49	4342.67	4772.44	4120.42	4116.68
	BIC		4370.20	4363.85	4802.75	4162.42	4192.29

Note. N = 166 after listwise deletion per variables in Model E. Model A = unconditional means model. Model B = unconditional linear growth. Model C adds a time varying predictor, sessions since last measurement. Model D includes hypothesized predictors. Model E = final model, including significant control variables and hypothesized and significant personality predictors. For pseudo R^2, Model B was compared to Model A, but subsequent models were compared to Model B. ~ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

therapy (number of sessions attended since the last measurement, as some clients came more than once a week, and others much less frequently). Although this led to accounting for 17% additional variance in the rate of change, the parameter estimate for the predictor was non-significant, and this variable was excluded in subsequent models.

Scores on Neuroticism and Extraversion were hypothesized to associate with the initial status of clients (indicating more and less psychological difficulty, respectively), and scores on Openness, Conscientious and Honesty/Propriety were hypothesized to predict increased benefit from therapy in terms of greater or faster decrease in OQ-45 scores. These predictors were added in Model D. Together, Neuroticism and Extraversion explained 45% of the variance in initial status, although this appeared to be driven almost entirely by Neuroticism. The three hypothesized personality predictors of slope appeared to together explain about 17% of the variation in rate of change, although only H/P had a marginally significant parameter estimate ($p < .10$).

The full, conditional model (E) included the hypothesized predictors, and all other personality scale and control variables that played a significant role in predicting change in scores. Variables excluded after assessing for significance included gender, education, sexual orientation, religion, ethnicity, type of therapy, and total sessions attended. The final model indicates that initial OQ-45 score (intercept) is strongly predicted by score on the personality scale Neuroticism (with which it had a .69 correlation, as shown in Table 4.8). Slope was predicted by Honesty/Propriety and Neuroticism (in both cases, higher scores at intake predicted more decrease over time), and marginally by age (such that for every year over 18 at time of intake, scores can be expected to decrease by .01 more each week). Compared to the baseline model, this model explained 1% more variance in within person variation in scores, 2% more variance in initial score on the OQ, and the same amount of variance in the rate of change. This constitutes a insignificant improvement on the more parsimonious Model D, and while AIC values decrease slightly from D to E (4120 to 4117), BIC values increase (4162 to 4192). (Excluding the non-significant predictors [Extraversion on intercept, Conscientiousness and Openness on intercept] led to highly similar results to Model E, and slightly higher AIC and BIC values.)

CHAPTER V
DISCUSSION AND CONCLUSION

Study 1

As a general measure of psychological functioning and change over time, the OQ-45 total score has criterion validity, and the practical need for such a measure is illustrated by its wide adoption and translation. The intended subscale structure, however, has not been supported by studies of convergent and discriminant validity or by confirmatory factor analyses. The current study went beyond previous analyses by considering alternative theoretically-driven models -- personality dimensions, and more or less elaborated models of types of problems. Because the OQ-45 is used over time during the course of therapy, the models were also tested in sets of responses after five or six therapy sessions, and after ten to twelve sessions.

The intended subscale structure of the OQ-45 was compared to alternative models derived from personality psychology and rational groupings of psychological problems, using EFA and CFA, in three random samples of intake session responses, and in samples of responses from two later time points. While neither the intended structure nor personality models had 'good fit' per standard guidelines (Hu & Bentler, 1999; Kline, 2011), several provided acceptable fit per *a priori* indices, better than those reported previously for the OQ-45 (e.g. Bludworth et al., 2010).

The personality models of OQ items had fit equal to or marginally better than that of the intended structure. This is notable considering that items were not written to capture personality differences or to adequately cover this domain. Some support is thus provided for the hypothesis that fundamental individual psychological differences captured by robust dimensions of personality influence responses to OQ-45 items more strongly than the intended factors. It has been well established that normal range temperament/personality attributes are closely related to specific psychological problems (e.g. Clark, 2005). For a clinician, a summary score of overall level of difficulty in functioning, answers to red flag items, and a profile of personality attributes might provide a more informative and parsimonious snapshot of a new client, and of change over time, than scores on the domains created by the OQ-45 authors. In addition to issues of structure, there is no evidence that the intended OQ-45 subscales aid mental health

treatment, but every reason to believe that knowledge of personality attributes could do so (Clark, 2009; Miller, 1991).

Best fit was observed for a seven-factor model of specific psychological problems, which was refined by observation of patterns of correlation in OQ-45 items. While this model had the advantage of additional factors, making its superiority over three- and four-factor models unremarkable, it is worth noting that this model evidenced good fit per standard CFA benchmarks (Hu & Bentler, 1999) in both an oblique factor and a bi-level version, despite being the product of an item-level analysis (Kline, 2011; Marsh et al., 2005) on a broad range inventory (Hopwood & Donnellan, 2010). The bi-level version has the strength of matching the theory and best-validated usage of the OQ-45 (including total score), while the standardized parameter estimates of the oblique model were more interpretable. This model, with items clustered using strong face- and empirical-validity, fit the observed data better than any previously reported model of the OQ-45. Moreover, the domains (depressive thinking, positive affectivity, anxiety, work/school adjustment, family/relationship stress, conflict, and substance abuse) are straightforward to interpret, map more closely onto DSM categories, and are more varied – they could prove more useful distinctions in empirical research and for practical use. Differential changes on subscale scores could be highly informative to clinicians. For example, anxiety and depression might increase after substance abuse decreases -- a natural step in the process of treatment, which could increase OQ-45 total score. In such a case, the specific subscale scores could confirm that therapy is proceeding in the right direction, despite the overall increase. Or a therapist might be able to see that while depression is decreasing, relationship difficulties have failed to change, and can use this information to guide the treatment plan.

In terms of the total score, the OQ-45 is not a unidimensional inventory. This does not seem to preclude it from functioning practically as a cumulative score of psychological problems, as it has demonstrated consistent predictive and convergent validity in terms of overall mental health functioning. In this sense, the inventory may work as a simple linear combination, and function outside the assumptions of Classical Test Theory in terms of internal consistency. This may be similar to the functioning of a measure of life stressors, for example, which estimates the amount of stress being

experienced by an individual, and may serve a practical function even if the items that are added together are not correlated. (For example, a divorce and a recent death in the family will likely have an additive stress effect without being correlated in the population.)

In addition to the red flag items, CFT supervisors report that they find the OQ-45 total score helpful for case review with students, because it clarifies whether the presenting complaint exists in a context of few or many psychological difficulties (T. Brown, personal communication, October 25, 2011). This is analogous to the way Miller (1991) advocates the use of Neuroticism scores in treatment – as a way to set appropriate expectations for the course of therapy, and perhaps to help a client come to terms with the texture of their experience. The OQ-45 total score could likely be improved psychometrically, to aid in this purpose, by either excluding reverse keyed items, or including an equal number of them. And results of the current study converge with those of other authors, in the recommendation that item 14 be removed from the OQ-45. It appears to be unassociated with other items, or to contradict them, and it lacks face validity as an item related to psychological difficulties. The OQ-45 would also be improved by clearer specification of what lower level domains are needed to complement the total score, and potentially by shortening the scale.

In summary, results provided tentative support for the hypothesis that normal-range dimensions of personality attributes may underlie scores on the OQ-45. Individual differences in basic patterns of psychological reactivity have been well mapped by personality psychologists, and these fundamental, temperamental tendencies are likely to influence responses on psychological inventories intending to capture a broad range of psychological phenomena.

Study 2

The goals of Study 2 were threefold: to test the models of OQ-45 structure identified in Study 1 in a non-clinical sample; to compare the OQ-45 to personality dimensions to explore the incremental validity of the OQ-45 over and above personality/temperament characteristics; and to select OQ-45 items for potential use as personality/temperament indicators.

In terms of the first goal, highly similar fit was observed in the student sample as compared to the clinical sample. In terms of the second, OQ-45 total score and Symptom Distress subscales correlated highly enough with QB6 Resiliency to suggest that these measures capture largely the same content. However, the six QB6 scales together only explain about half the variance in OQ total scores. This suggests one of two possibilities. The OQ may function as intended, capturing state variation and nuances in level of psychological problems that go beyond what is captured by a normal-range personality inventory. Or, the OQ-45's ability to capture overall psychological functioning could be explained by this covariation with a well-validated model of individual differences, and the unexplained variance could be due to measurement error. A comparative validity analysis that tests the relative abilities of a personality inventory and the OQ-45 to predict meaningful differences in presentation and treatment outcomes could shed light on whether the OQ-45 tells us something more than a normal-range personality questionnaire (plus risk factor items) could tell us. (Addressed in Study 3).

Finally, OQ-45 items useable as markers of personality attributes were identified. While these subcomponents do not comprise the full range of individual psychological differences encompassed by a broad inventory like the QB6, they do capture aspects of personality relevant to clinical treatment. These scales may allow researchers working with OQ-45 datasets to make preliminary inferences about the personality characteristics of clients, allowing for links to the well-established literature in personality psychology.

Study 3

Study 3 explored how robust personality constructs might be used to better guide treatment decisions and assess treatment efficacy. Self-report scores on Big Five personality trait dimensions in adulthood relate to virtually all measurable life outcomes, and have been reliably associated with clinical presentation and disorders (e.g. Clark, 2005). Less research has addressed how personality traits affect and moderate treatment in psychotherapy. Study 3 contributed to the literature by testing the relation of Big Six dimensions to psychological functioning at presentation, to usage of therapy, to termination outcome, and to change in psychological functioning, in a sample of community clinic clients with a diverse range of presenting concerns. To address questions raised in Studies 1 and 2, the comparative validity of the OQ-45 and the Big

Six to predict clinically relevant behavior and outcomes was tested using the outcomes unrelated to OQ-45 scores (sessions attended and therapy outcome).

It was hypothesized that scores on Neuroticism and Extraversion, and to a lesser extent those on Conscientiousness, Agreeableness, and Honesty/Propriety, would associate with psychological functioning at intake. Neuroticism in particular is associated with clinical disorders (Clark, 2005), and with the usage of therapy (Miller, 1991). The association between Neuroticism and the OQ-45 was indeed very strong ($r = .69$), although not to the extent that the measures should be assumed to be synonymous with one another. The other scales were also meaningfully associated with the OQ-45, with correlations ranging from .22 to .41, with the anticipated exception of Openness.

In terms of sessions attended, it was hypothesized that Conscientiousness, Honesty/Propriety, and Agreeableness would play significant roles. Conscientiousness is believed to predict the ability to benefit from therapy by diligently adhering to interventions (Miller, 1991). Honesty/Propriety was hypothesized to predict motivation for self-improvement and taking responsibility for one's own shortcomings, and thus effective usage of therapy; those with very low levels are predicted to engage in therapy only when mandated or for secondary gain. Agreeableness has been shown to relate to the formation of therapeutic alliance (Coleman, 2006), but is also thought to potentially inhibit necessary self-assertion in therapy (Miller, 1991). For this reason, Conscientiousness and Honesty/Propriety (but not Agreeableness) were hypothesized to predict more successful termination outcome. Both Extraversion and Openness have been associated with alliance formation (Coleman, 2006) and both are thought to relate to interest in engaging in therapy (Miller, 1991), although in neither case was this expected to be a large effect in the current study.

Results indicated no association between any personality scales and the number of sessions attended or termination outcome rating. Instead, only age, level of education, being married, and attending two types of therapy at the clinic predicted attending more sessions. And only level of education and income, and being female, predicted more successful outcomes (per therapist/researcher rating).

In terms of change in scores on the OQ-45, Conscientiousness, Honesty/Propriety, and Openness were predicted to predict steeper slope. Openness was added here because

it is thought that people high in this dimension might be more flexible and able to change, and because of the association with interest in therapy. The hypothesis for Honesty/Propriety was supported, in that those with higher scale scores at intake had greater decrease in OQ-45 scores over time, but there was no significant relation for Conscientiousness or Openness. Neuroticism scores also predicted greater decrease over time. The only demographic predictor of slope, of marginal significance and very small effect, was age, in that older people saw slightly less decrease per week.

The role of Honesty/Propriety in benefiting from therapy suggests the importance of the addition of this construct to a comprehensive model of personality attributes. This QB6 dimension (Saucier, 2009), which was added to the regular Big Five Inventory as in Thalmayer et al. (2011), is similar to the Honesty/Humility factor in the HEXACO six-factor model (Ashton & Lee, 2007). This content, related to ethical behavior and integrity versus greed, tends to be left out of the Big Five model, but has consistently appeared across diverse lexical studies (Ashton & Lee, 2010; Saucier, 2009). This content logically relates to benefiting from mental health treatment, because it indicates the personal integrity that could lead to taking responsibility for personal problems and attempting to overcome them. As noted above, scores on this domain have been shown to correlate more highly than other Big Five or Six dimensions with compulsive drinking, risk taking, and lawbreaking behaviors (Saucier, 2009), and HEXACO Honesty/Humility has demonstrated strong negative correlations with measures of the 'dark triad': psychopathy, Machiavellianism and narcissism (Lee & Ashton, 2005). Given that these conditions are notoriously difficult to treat, it is logical that persons high in Honesty/Propriety should be more "treatable", and should see more benefit over time from therapy, than those with low scores in this domain.

The role of Neuroticism in benefitting from therapy may be partly due to regression to the mean -- higher scores at intake were associated with significantly higher intake OQ-45 scores. It also may suggest that some component of Neuroticism is sensitive to mental health treatment. De Fruyt and colleagues (2006) note that while personality continuity is a precondition for using scores to predict disorders, studies often show decreases in Neuroticism after treatment.

Taken together, the results of the current study do not provide strong support for the overall hypothesis that knowledge of personality attributes could meaningfully guide therapists in their treatment efforts. This may partly be due to lack of power. The sample size is small, given the small to medium effects generally expected with short, self-report personality inventories. In general, the strength of such inventories is their ease of use and reliability, not their large effect sizes. For this reason, samples of 200 to 300 are preferable, but the samples usable in each of the current analyses hovered nearer 150.

Secondly, the context may not have been ideal for identifying such effects. The Couple and Family Therapy clinic is an advantageous research setting in several respects – the client population is highly diverse in terms of presenting complaint (and thus in personality characteristics), the clinic is busy, with several intakes per week, and the clinic keeps excellent records and collects extensive demographic information and multiple OQ-45 responses from clients. Furthermore, the site offered excellent ecological validity. It is surely more defensible to generalize from a community-serving clinic such as this one, than from the more selective situations, such as university counseling centers, inpatient hospitals, or controlled trials, from which data on therapy results is more often available. On the other hand, much control over extraneous variables and power to find effects is lost. The diversity of client complaints creates noise in the data that may obscure real associations. Some clients come in with highly complex issues and serious difficulties in functioning, while others arrive without clinically significant problems, simply seeking support through transitions or to iron out family conflicts.

An inhibiting factor specific to predicting change in OQ-45 with personality scales in this sample was the small amount of baseline change in scores. In the unconditional growth model, the average decrease in OQ-45 score per week was .26, a very small amount, considering the average intercept of 71.04. This estimate is similar to that observed in a much larger sample of clients from this clinic (Thalmayer & Baune, 2013). At this rate it would take clients about one year to achieve what the OQ-45 manual describes as "clinically significant change" – a 14 point decrease (Lambert, Morton, et al., 2004). While this is not a totally unreasonable timeline, it would demand longer term therapy than is most typically offered to clients at the CFT. (Mean sessions attended for the current sample was less than eight.) On the other hand, about 40% of the sample

analyzed had intake OQ-45 scores below the cutoff for clinical significance (63; Lambert, Morton, et al., 2004). For such clients, it is likely both unnecessary, and unrealistic, to see a large decrease in scores, and there may be a ceiling effect on change. Running the unconditional growth model including only clients with intake OQ total of 63 or higher does lead to a slightly steeper slope (-.36 per week), though the sample size is small and other results do not appear to differ from those reported. In a larger sample, further analyses could more thoroughly explore differences for different types of clients.

There are other possible explanations for the small amount of average change. One is that this accurately reflects the modest levels of change typically achieved during brief, solution-focused therapy. This predominantly low-income clientele may have approached therapy with practical goals in mind, and neither sought, nor were offered, services focused on deep, long term change. And, while the CFT is a highly regarded master's level training program, the therapists are in their first or second years of treating clients, and may not yet have the experience to help clients achieve such significant changes. Finally, change that did occur might not have been captured by the timing of measurement. Because the OQ-45 is administered in this clinic for the practical purpose of guiding treatment, it is administered more frequently at the beginning of therapy than toward the end. There is no policy of obtaining a score at termination, let alone post-therapy, which could have allowed for more accurate estimation of final OQ-45 status.

The comparative validity of the Big Six personality inventory and the OQ-45 to predict clinical outcomes was tested to address questions raised in Studies 1 and 2. The results were largely a draw – no scales of either inventory significantly predicted termination outcome, and variance explained in outcome was the same for both models. The OQ-45 subscale model, however, bested the Big Six in predicting sessions attended, due to a significant association with the Interpersonal Relations scale. This finding is interesting given the family-systems focus of the clinic, though it is hard to interpret why clients with interpersonal complaints should have attended more sessions without being more likely to terminate successfully. While attending sessions is a necessary prerequisite for successfully meeting therapeutic goals, it is not a sufficient one.

Aside from providing evidence for the role of Honesty/Propriety in predicting improved psychological functioning during therapy, the current results do not provide

strong support for the hypothesis that knowledge of personality attributes could guide therapists in their treatment efforts. Despite the limitations and limited findings of the current study, it remains likely that personality attributes play a role in therapy usage and outcomes, as they strongly associate with clinical presentation and with diverse life outcomes. Future studies might benefit from a larger sample size, where it is possible to systematically separate clients by type of therapy, intake OQ scores, and other factors. It would also be beneficial to collect data in other types of clinics. In particular, psychologists in private practice or clinics that can offer a diversity of treatment styles may most benefit from knowledge of personality (as in Miller, 1991), and it might be most suitable to test the role of personality attributes in such settings.

Future studies could also address how traits change during therapy, and whether such change should be considered a treatment goal. Personality attributes are known to be largely but not entirely stable (Roberts, Walton, & Viechtbauer, 2006). Understanding how attributes play out in therapy could lead to addressing such tendencies within therapy. This could help clinicians address client's problems in living that go beyond meeting criteria for a disorder (Clark, 2009).

Conclusion

The purpose of this dissertation was to explore ways that the efficient and reliable measurement of normal-range psychological differences might improve the administration of mental health treatment. The specific aims were twofold: To explore the question of optimal outcome measurement in psychotherapy, and to determine how individual differences in personality attributes affect the course of therapy.

In Study 1 the structure of the OQ-45, a popular measure designed to capture change in broad functioning, was compared to theoretically determined alternative structural models in a large sample of community clinic clients. In Study 2 the alternative structures were tested in a college student sample, and relations with Big Six personality attribute dimensions were explored. The fit of the models across samples was similar. A four-factor model of personality dimensions fit the data slightly better than the intended subscale structure, providing tentative support for the hypothesis that normal-range dimensions of personality attributes underlie scores on the OQ-45. Robust individual differences in basic patterns of psychological reactivity have been well mapped by

personality psychologists. Such fundamental, temperamental tendencies are likely to influence responses on psychological inventories intending to capture a broad range of psychological phenomena. A seven-factor model of problems, however, provided the best fit, and may have the most to offer clinicians.

Study 2 also explored the relation of the OQ-45 total score and subscales of the intended and alternative models to Big Six personality attribute scales. As hypothesized, Resiliency (vs. Neuroticism) correlated most highly with OQ total score and the Symptom Distress subscale. Together the six QB6 scales explain about half the variance in OQ total score. OQ-45 items that can be used as markers for subcomponents of personality attributes were identified.

Study 3 assessed how individual differences in personality attributes relate to psychological functioning at presentation, and how they predict therapy usage, termination outcome, and change in psychological functioning, in a sample of community clinic clients with diverse presenting concerns. Of the personality variables, only Honesty/Propriety and Neuroticism were shown to play a role in mental health treatment, in that scores on these dimensions predicted more decrease in OQ-45 scores over time.

The comparative validity of the OQ-45 versus the Big Six personality model to predict clinically relevant behavior and outcomes was tested using the two outcomes unrelated to OQ-45 scores, sessions attended and therapy outcome. While the inventories both failed to predict termination outcome, the interpersonal relations subscale of the OQ-45 predicted attending more sessions.

The results of the current studies may help clinics decide whether the OQ-45 is the ideal instrument for their purposes. Published reports of its reliability, validity, and typical levels of change do not yet exist for couples and family therapy clinics using the systems model of family therapy. It may be the other items and scales would better capture the kind of improvement achieved in such settings. The measure is also longer than is likely necessary, and probably too long to obtain valid scores from people with lower education, or those in crisis. A briefer measure with stronger psychometric properties, which could be administered more frequently, and at termination, might help such clinics better capture the results of treatment, and better allow for the identification of predictors of change. On the other hand, CFT supervisors report that they find total

score reports and red flag items helpful in case review and helping students set priorities (T. Brown, personal communication, October 25, 2011). Such clinics may be getting acceptable value from the OQ-45, even without making use of subscale scores or tracking clinically significant change.

Most importantly, my goal was to find ways to make therapy successful for a more diverse range of people. Many treatments exist, but drop out rates are high, and success rates are moderate. Low efficiency contributes to costs, a serious factor in a country striving to increase access to health services. There is a substantial literature on normal range personality variation, but very few extensions of this knowledge base into clinical practice have been made. Future work may establish more clearly the role that personality attributes play. It remains important to find ways to best match treatment modalities and therapists to clients, and to establish appropriate treatment modifications for clients who do not or cannot benefit from standard approaches. Therapy retention, efficiency, and success rates need to be improved, and continued exploration of the appropriate use of measurement may help determine ways to do so with minimal cost and effort.

APPENDIX

ALTERNATIVE MODELS OF OUTCOME QUESTIONNAIRE-45 ITEMS

The Appendix Table shows the items anticipated to load on each subscale. The scales for the alternative models were initially rationally constructed using the face validity of items. For personality scales, choices were driven by consideration of Big Five (John & Srivastava, 1999) and Big Six (Ashton & Lee, 2007; Saucier, 2009) models of adult personality and three- and four- factor models of childhood temperament (Clark, 2005; Rothbart, 2007), and by the available content in OQ-45 items. Gerard Saucier, an expert on personality structure and measurement, was consulted. Although models of personality structure including five or six dimensions provide more comprehensive coverage of the adult personality variation, the OQ-45 does not include items that relate to all Big Five and Big Six domains – in particular, no items appear to relate to the Openness/Intellect/Originality domain. For reasons of availability, three- and four-factor models of personality/temperament structure were developed. The names of the scales were chosen to best capture the dominant content.

Table A.1. Alternative models of Outcome Questionnaire-45 item structure.

Scales	OQ Items
Intended Structure	
Symptom Distress	2, 3, 5, 6, 8, 9, 10, 11, 13R, 15, 22, 23, 24R, 25, 27, 29, 31R, 33, 34, 35, 36, 40, 41, 42, 45
Interpersonal Relations	1R, 7, 16, 17, 18, 19, 20R, 26, 30, 37R, 43R
Social Role	4, 12R, 14, 21R, 28, 32, 38, 39, 44
Three factor personality model*	
Negative Affectivity	2, 4, 5, 6, 8, 9, 10, 15, 16, 18, 22, 23, 25, 27, 28, 29, 33, 34, 35, 36, 38, 40, 41, 42, 45
Positive Affectivity	3, 7, 12R, 13R, 17, 20R, 21R, 24R, 31R, 37R, 43R
Disinhibition	1R, 11, 14, 19, 26, 30, 32, 39, 44
Four-factor personality model	
Negative Affectivity	2, 4, 5, 6, 8, 9, 10, 15, 22, 23, 25, 27, 28, 29, 33, 34, 35, 36, 38, 40, 41, 42, 45
Positive Affectivity	3, 2R, 13R, 21R, 24R, 31R
Disinhibition	11, 14, 26, 32, 39, 44
Affiliation	1R, 7, 16, 17, 18, 19, 20R, 30, 37R, 43R

Table A.1. (continued)

Scales	OQ Items
Internalizing/Externalizing Tendencies**	
Internalizing	2, 3, 4, 5, 6, 7, 8, 9, 10, 12R, 13R, 15, 16, 18, 20R, 21R, 22, 23, 24R, 27, 28, 29, 31R, 33, 34, 35, 36, 37R, 38, 40, 41, 42, 43R, 45,
Externalizing	1R, 11, 19, 26, 30, 32, 39, 44
Seven-factor Psychological Problems Model	
Depressive Thinking	3, 5, 8, 9, 10, 15, 23, 40, 42
Positive Emotionality	12R, 13R, 21R, 24R, 31R, 43R
Anxiety	25, 27, 29, 33, 34, 35, 36, 41, 45
Work/School Adjustment	2, 4, 22, 28, 38
Family/Relationship Stress	7, 16, 17, 18, 19, 20R, 37R
Conflict	1R, 6, 30, 39, 44
Substance Abuse	11, 26, 32

Note. IS = Intended structure with subscales SD = Symptom Distress, IR = Interpersonal Relations, SR = Social Role. 3fP = Three-factor personality model with subscales NA = Negative Affectivity, PA = Positive Affectivity, D = Disinhibition; items moved after EFA before CFA: 1 from PA to D; 14 included on D for direct comparison of the three factor models; 17 from unaffiliated to PA. 4fP = Four-factor personality model – AF = Affiliation. Modification made after initial CFA: item 14 was removed from the model (previously on Disinhibition), I/E = internalizing/externalizing model of psychological problems. 7 = Seven-factor model of psychological problems includes subscales of: DT = Depressive Thinking, PA = Positive Affectivity, Anx = Anxiety, WSA = Work/School Adjustment, FRS = Family/Relationship Stress, Con = Conflict, SA =Substance Abuse.

* The three substance use items include the instructions: (If not applicable, mark 'Never')

Initial subscale placement for the internalizing externalizing model was determined with reference to Kreuger and Markon (2006). Scale placement for the seven-factor problems model was determined rationally with reference to DSM distinctions.

After the initial exploratory factor analyses, some models were updated, as detailed in the table – logical alterations were made to the 3-factor personality model, the four-factor personality model and the seven-factor problems model.

REFERENCES CITED

American Psychiatric Associations (2000). *Diagnostic and statistical manual of mental disorders.* Text revision (4th ed., rev. ed.). Washington, DC: Author.

Anderson, K. W, & McLean, P. D. (1997). Conscientiousness in depression: Tendencies, predictive utility, and longitudinal stability. *Cognitive Therapy and Research, 2,* 223-238.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11,* 150 –166. doi:10.1177/1088868306294907

Ashton, M., & Lee, K. (2010). On the cross-language replicability of personality factors. *Journal of Research in Personality, 44,* 436–441.doi:10.1016/j.jrp.2010.05.006

von Bergen, A., & de la Parra, G. (2002). OQ-45.2, an Outcome Questionnaire for monitoring change in psychotherapy: Adaptation, validation, and indications for its application and interpretation. *Terapia Psicológica, 20,* 161-176.

de Beurs, E., den Hollander-Gijsman, M., Buwalda, V., Trijsburg, W., & Zitman, F. (2005). The Outcome Questionnaire (OQ-45): Measuring psychiatric symptoms and interpersonal functioning. *Psycholoog, 40,* 393-400.

Bludworth, J. L., Tracey, T. J. G., & Glidden-Tracey, C. (2010). The bilevel structure of the Outcome Questionnaire-45. *Psychological Assessment, 22,* 350-355.

Bodlund, O., Kullgren, G., Ekselius, L., Lindström, E., & von Knorring, L. (1994). Axis V— Global Assessment of Functioning Scale: Evaluation of a self-report version. *Acta Psychiatrica Scandinavica, 90,* 342-347.

Bouchard, T. J (2004). Genetic influence on human psychological traits. *Current Directions in Psychological Science, 15*, 148-151.

Carver, C.S. (2005). Impulse and constraint: Perspectives from personality psychology, convergence with theory in other areas, and potential for integration. *Personality and Social Psychology Review, 9*, 312-333.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A. & Poulton, R. (2003). Effect of life stress on depression moderated by 5-HTT gene. *Science, 301,* 386-389.

Claridge G. & Davis, C. (2001). What is the use of Neuroticism? *Personality and Individual Differences, 31,* 383-400.

Clark**,** L. A**.** (2005). Temperament as a unifying basis for personality and psychopathology. *Journal of Abnormal Psychology, 114, 505-521.*

Clark, L. A. (2009). Stability and change in personality disorder. *Psychological Science, 18,* 27-32.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7,* 309–319. doi:10.1037/1040-3590.7.3.309

Clark, L. A., Watson, D., & Mineka, S. (1994). Temperament, personality, and the mood and anxiety disorders. *Journal of Abnormal Psychology, 103,* 103–116.

Coco, G. L., Chiappelli, M., Luca, B., Gullo, S., Prestano, C., & Lambert, M. J. (2008) The factorial structure of the Outcome Questionnaire-45: A study with an Italian sample. *Clinical Psychology & Psychotherapy, 15*, 418-423.

Coker, L. A. & Widiger, T. A. (2005). Personality disorders. In Maddux, J. E., & Winstead, B. A. (Eds.), *Psychopathology: Foundations for a contemporary understanding* (pp. 201-227). Mahwah, NJ: Erlbaum.

Coleman, D. (2006). Client personality, working alliance and outcome: A pilot study. *Social Work in Mental Health, 4,* 83-98. doi: 10.1300/J200v04n04_06.

Costa, P. T. & McCrae, R. R. (1989) *Revised NEO Personality Inventory (NEO-PI-R)*. Psychological Assessment Resources, Inc.

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2005). Sources of Openness/Intellect: Cognitive and neuropsychological correlates of the fifth factor of personality. *Journal of Personality, 73,* 825-858.

Dishion, T. J. & Stormshak, E.A. (2007) Intervening in children's lives: An ecological, family-centered approach to mental health care. Washington D.C.: American Psychological Association.

Doerfler, L. A., Addis, M. E., & Moran, P. W. (2002). Evaluating mental health outcomes in an inpatient setting: Convergent and divergent validity of the OQ-45 and BASIS-32. *The Journal of Behavioral Health Services & Research, 29*, 394-403.

De Fruyt, F., Van Leeuwen, K., Bagby, R. M., Rolland, J., & Rouillon, F. (2006). Assessing and interpreting personality change and continuity in patients treated for major depression. *Psychological Assessment,18,* 71-80.

Egan, J. (2005). Dropout and related factors in therapy. *The Irish Psychologist, 32,* 27-30.

Feder, L. & Dugan, L. (2004). Testing a Court-Mandated Treatment Program for Domestic Violence Offenders: The Broward Experiment. Report submitted to National Institute of Justice, Office of Justice Programs, U.S. Department of Justice.

Funder, D. (2007). *The Personality Puzzle*. New York, NY: Norton & Co.

Haug, S., Puschner, B., Lambert, M. J., & Kordy, H. (2004). Assessment of change in psychotherapy with the German version of the Outcome Questionnaire (OQ-45.2). *Zeitschrift für Differentielle und Diagnostische Psychologie, 25,* 141-151.

Hess, T. R., Rohlfing, J. E., Hardy, A. O., Glidden-Tracey, C., & Tracey, T. J. G. (2010) An examination of the "interpersonalness" of the Outcome Questionnaire. *Assessment, 17,* 396-399.

Hilsenroth, M. J., Ackerman, S. J., Blagys, M. D., Baumann, B. D., Baity, M. R., Smith, S. R., Price, J. L., & Holdwick, D. J. J. (2000). Reliability and validity of DSM-IV Axis V. *The American Journal of Psychiatry, 157,* 1858-1863.

Hogan, R. (1983). A socioanalytic theory of personality. *Nebraska Symposium on Motivation,* 55-89.

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review, 14*, 332-346.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory* (Versions 4a and 54). Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, O. P. & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In Pervin, L. A. & John, O. P. (Eds.), *Handbook of Personality: Theory and Research.* NY: The Guilford Press.

de Jong, K., Nugter, M. A., Polak, M. G., Wagenborg, J. E. A., Spinhoven, P., & Heiser, W. J. (2007). The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation. *Clinical Psychology & Psychotherapy, 14,* 288-301.

Jurado, J. (2007). Towards an understanding of the psychometric properties of the Spanish version of the OQ-45. *Dissertation Abstracts International: Section B: The Sciences and Engineering,* p. 3400.

Kim, S., Beretvas, S. N., & S., & Sherry, A. R. (2010) A validation of the factor structure of OQ-45 scores using factor mixture modeling. *Measurement and Evaluation in Counseling and Development, 42,* 275-295.

Kline, R. B. (2011). Principles and practice of structural equation modeling. New York, NY: Guilford Press.

Krueger, R.F. & Markon, K. E. (2006). Reinterpreting comorbidity: A model–based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology, 2,* 111-133.

Lambert, M. J., Gregersen, A. T. & Burlingame, G. M., (2004). The Outcome Questionnaire-45. In Maruish, M. E. (Ed.), *The use of psychological testing for treatment planning and outcomes assessment: Volume 3: Instruments for adults* (3rd ed., pp. 191-234). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Lambert, M. J., Morton, J.J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., Shimokawa, K. Christopherson, C., & Burlingame, G. M. (2004). *Administration and scoring manual for the OQ-45.2 Outcome Questionnaire.* Salt Lake City, UT: American Professional Credentialing Services, L.L.C.

Lambert, M. J., Smart, D. W., Campbell, M. P., Hawkins, E. J., Harmon, C., & Slade, K. L. (2006). Psychotherapy outcome, as measured by the OQ-45, in African American, Asian/Pacific Islander, Latino/a, and Native American clients compared with matched Caucasian clients. *Journal of College Student Psychotherapy, 20,* 17-29.

Lambert, M. J., & Vermeersch, D. A., (2008). Measuring and improving psychotherapy outcome in routine practice. In Brown, S. D. & Lent, R. W. (Eds), *Handbook of counseling psychology* (4th ed., pp. 233-248). Hoboken, NJ: John Wiley & Sons Inc.

Lee, K., & Ashton, M. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*, 329–358.

Lee, K., & Ashton, M. C. (2005). Psychopathy, Machiavellianism, and Narcissism in the Five-Factor Model and the HEXACO model of personality structure. *Personality and Individual Differences, 38,* 1571-1582.

Li, Y. & Luo, H. (2009). The reliability and validity of the Outcome Questionnaire-Chinese version. *Chinese Mental Health Journal, 23,* 105-107.

Lilenfeld, L. R. R., Wonderlich, S., Riso, L. P., Crosby, R., & Mitchell, J. (2006). Eating disorders and personality: A methodological and empirical review. *Clinical Psychology Review, 26,* 299-320.

Lowe, J. R. & Widiger T. A. (2008). Personality disorders. In Maddux, J. E., & Winstead, B. A. (Eds), *Psychopathology: Foundations for a contemporary understanding.* Mahwah, NJ: Erlbaum.

de Manzano, O., Cervenka, S., Karabanov, A., Farde, L., & Ullén, F. (2010). Thinking outside a less intact box: Thalamic dopamine D2 receptor densities are negatively related to psychometric creativity in healthy individuals. *Plos One, 5,* 1-6.

Marsh, H. W., Hau, K., & Grayson, D. (2005) Goodness of fit in structural equation models. In Maydeu-Olivares, A., & McArdle, J. J. (Eds) *Contemporary psychometrics*. Mahwah, N.J: Lawrence Erlbaum.

Masi, M. V., Miller, R. B., & Olson, M. M. (2003). Differences in dropout rates among individual, couple, and family therapy clients. *Contemporary Family Therapy, 25,* 63-75.

May, R. (1967). *Psychology and the human dilemma.* Princeton, N.J., Van Nostrand.

McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology, 52,* 1258-1265.

Miller, T. R. (1991). The psychotherapeutic utility of the Five-Factor Model of personality: A clinician's experience. *Journal of Personality Assessment, 57,* 415-433.

Minami, T., Davies, D. R., Tierney, S. C., Bettmann, J. E., McAward, S. M., Averill, L. … Wampold, B. E. (2009). Preliminary evidence on the effectiveness of psychological treatments delivered at a university counseling center. *Journal of Counseling Psychology, 56,* 309-320.

Minuchin, S. (1974). *Families and family therapy*. Cambridge, MA: Harvard University Press.

Moutafi, J., Furnham, A., & Crump, J. (2006). What facets of openness and conscientiousness predict fluid intelligence score? *Learning and Individual Differences, 16*, 31-42.

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

Mulder, R. T. (2002). Personality pathology and treatment outcome in major depression: A review. *The American Journal of Psychiatry.159,* 359-371.

Nettle, D. (2007). *Personality: What makes you the way you are.* Oxford University Press.

Nielsen, S. L., Lambert, M. J., Smart, D. W., Okiishi, J., Isakson, R. L., Pedersen, T., & Nielsen, D.L. (2010). The implementation of empirically supported psychotherapy: Practice foundations. In Carlstedt, R. A. (Ed), *Handbook of integrative clinical psychology, psychiatry, and behavioral medicine: Perspectives, practices, and research* (pp. 59-78). New York, NY: Springer Publishing Co.

Oltmanns & Turkheimer, E. (2009). Personality pathology. *Psychological Science, 18,* 32-37.

Orlinksy, D. E. (2010). Foreword. In Duncan, B.L., Miller, S. D., Wampold, B.E., & Hubble, M. A. (Eds.) *The heart and soul of change: Delivering what works in therapy* (2nd. ed., pp. xix-xxv). Washington, D.C.: American Psychological Association.

Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57,* 401-21.

Qin, Y., Hu, S. (2008). Usability report of Outcome Questionnaire-45 in part of Chinese sample *Chinese Journal of Clinical Psychology, 16,* 138-140.

Quilty, L. C., De Fruyt, F., Rolland, J., Kennedy, S. H., Rouillon, F., & Bagby, R. M. (2008) Dimensional personality traits and treatment outcome in patients with major depressive disorder. *Journal of Affective Disorders, 108*, 241–250.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41,* 203–212.

Roberts, B. R. Walton, K. E. & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132,* 1–25.

Rothbart, M.K. (2007). Temperament, development, and personality. *Current Directions in Psychological Science, 16,* 207-212.

Samuel, D. B., & Widiger, T. A. (2006). Clinicians' judgments of clinical utility: A comparison of the DSM-IV and Five-Factor models. *Journal of Abnormal Psychology, 115,* 298-308.

Saucier, G. (2009). Recurrent personality dimensions in inclusive lexical studies: Indications for a Big Six structure. *Journal of Personality, 77*, 1577-1614.

Saucier, G., Thalmayer, A.G., & Bel-Bahar, T. (2013). Personality descriptors ubiquitous across 12 languages. Manuscript submitted for publication.

Saucier, G., Thalmayer, A. G., Payne, D. L., Carlson, R., Sanogo, L., Ole-Kotikash, L., …. & Zhou, X. (2013). A basic bivariate structure of personality attributes evident across nine languages. *Journal of Research in Personality,* in press.

Shafran, R., Clark, D. M., Ehlers, A., Garety, P. A., Salkovskis, P. M., Fairburn, C. G., Williams, J. M. G., & Wilson, G. T. (2009). Mind the gap: Improving the dissemination of CBT. *Behaviour Research and Therapy, 47,* 902-909.

Singer, J. D., & Willett, J. B. (2003). Applied longitudinal data analysis: Modeling change and event occurrence. Oxford University Press.

Slutske, W.S., Caspi, A., Moffitt, T. E., & Poulton, R. (2005). Personality and problem gambling: A prospective study of a birth cohort of young adults. *Archives of General Psychiatry, 62,* 769-775.

Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology, 84,* 1041–1053.

Soldz, S., & Vaillant, G. (1999). The Big Five personality traits and the life course: A 45-year longitudinal study. *Journal of Research in Personality, 33,* 208.

Sue, S. (2006). Cultural competency: From philosophy to research and practice. *Journal of Community Psychology, 34,* 237-245.

Talley, J. E., & Clack, R. J. (2006). Use of the Outcome Questionnaire 45.2 with a university population. *Journal of College Student Psychotherapy, 20,* 5-15.

Thalmayer, A. G. & Baune, N. (2013). Client predictors of therapy usage and outcome. Unpublished manuscript.

Thalmayer, A. G. & Saucier, G. (2011). Mature values, personality and psychological change. Poster presented to the Association for Research in Personality annual meeting, Riverside, CA, June 2011.

Thalmayer, A. G. Saucier, G. & Eigenhuis, A. (2011). The comparative validity of brief-to medium–length big five and big six personality inventories. *Psychological Assessment, 23,* 995-1009.

Trull T. J. & Sher, K. J. (1994). Relationship between the five-factor model of personality and Axis I disorders in a nonclinical sample. *Journal of Abnormal Psychology, 103*, 350-360.

Umphress, V. J., Lambert, M. J., Smart, D. W., Barlow, S. H., & Clouse, G. (1997). Concurrent and construct validity of the outcome questionnaire. *Journal of Psychoeducational Assessment,* 15, 40-55.

Vermeersch, D. A., Whipple, J. L., Lambert, M. J., Hawkins, E. J., Burchfield, C. M. & Okiishi, J. C. (2004). Outcome Questionnaire: Is it sensitive to changes in counseling center clients? *Journal of Counseling Psychology, 51,* 38-49.

Wennberg, P., Philips B., & De Jong, K. (2010) The Swedish version of the Outcome Questionnaire (OQ-45): Reliability and factor structure in a substance abuse sample. *Psychology and Psychotherapy: Theory, Research and Practice, 83*, 325-329.

Westen, D., & Bradley, R. (2005). Empirically supported complexity. *Psychological Bulletin, 14,* 266-271.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin, 130,* 631-663.

Whittle, S., Allen, N. B., Lubman, D.I., & Yücel, M. (2006). The neurobiological basis of temperament: Towards a better understanding of psychopathology. *Neuroscience & Biobehavioral Reviews, 30*, 511-525.

Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist, 62,* 71-83.

Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice, 24,* 190-195.