



Theses and Dissertations

2012-05-18

Measuring Growth: The Reliability and Validity of the Utah Recovery Scale

Ray J. Katzenbach
Brigham Young University - Provo

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Counseling Psychology Commons](#), and the [Special Education and Teaching Commons](#)

BYU ScholarsArchive Citation

Katzenbach, Ray J., "Measuring Growth: The Reliability and Validity of the Utah Recovery Scale" (2012).
Theses and Dissertations. 3561.
<https://scholarsarchive.byu.edu/etd/3561>

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Measuring Growth: The Reliability and Validity of the Utah Recovery Scale

R. Jason Katzenbach

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree

Doctor of Philosophy

Robert Gleave, Chair
Gary Burlingame
Mark Beecher
Lane Fischer
John Okiishi
Dallas Jensen

Department of Counseling and Special Education

Brigham Young University

August 2012

Copyright © 2012 R. Jason Katzenbach

All Rights Reserved

ABSTRACT

Measuring Growth: The Reliability and Validity of the Utah Recovery Scale

R. Jason Katzenbach

Department of Counseling Psychology & Special Education, BYU

Doctor of Philosophy

Recently the direction of consumer mental health care in the United States has shifted in terms of its approach to recovery. In this sense recovery is not thought to be a complete amelioration of symptoms, but rather the acquisition of meaningful relationships, independent living, and fulfilling work. In response to these changes, the Utah division of the National Alliance for the Mentally Ill (NAMI-Utah) conducted consumer focus groups for the purpose of developing a tool to monitor this new conceptualization of recovery. The focus groups generated 10 recovery indicators based on recovery as the Substance Abuse and Mental Health Services Administration have defined it. This study explored initial psychometric reliability and validity estimates for these recovery indicators and their ability to track changes in recovery over time. In addition, the study also explored the relationship between distress reduction and recovery both concurrently and over time.

Keywords: Recovery, Positive Psychology, Outcome, Managed Care, Utah Recovery Scale

ACKNOWLEDGEMENTS

First off, I would like to thank Nathan Hyer, Clint & Michelle Beebe, Christopher Cannon, and Megan Haskett Armstrong, for helping to keep me sane throughout graduate school and through some of the difficult personal challenges that I faced that almost prevented me from being able to complete this document.

I would also like to thank Dr. Liz Baker and Dr. Chris Chapman who gave me my first research opportunities as an undergraduate. In addition, I would like to thank Dr. Robert Gleave and Dr. Gary Burlingame for making me a part of their research team during my undergraduate education and for having faith enough in me to continue to trust me with interesting projects, as well as for reading several drafts of my work and for always giving me helpful feedback and encouragement. I would particularly like to thank Robert who has been my chair and who has been supportive of me throughout this process and throughout graduate school.

I would also like to thank Dr. John Okiishi whose class made it possible for me to get into graduate school; Dr. Dallas Jensen who has been a good friend and mentor to me on this project; Jon Katzenbach whose financial support made both my undergraduate and graduate work possible; Dr. Jenith Larsen for having enough faith in me to work with me on research while I was an undergraduate; and Bob Kellet for giving me a campus job that allowed me to acquire the necessary computer skills that I needed in order to complete the data analysis for this project.

I would also like to thank my graduate cohort: Kristina Hansen, Kirti Potkar, Candilyn Newell, and Leslie Clark. They were always supportive throughout my training and always graciously put up with my heretical ranting. I would also like to thank the Counseling Psychology Special Education faculty for pushing me to think deeper and to be a better

professional. I would also like to thank the BYU Counseling and Career Center faculty that graciously accepted me as a colleague and helped me to develop both as a professional and as a therapist. Lastly, I would also like to thank Arna Clark and Judy Macallister at Alta High School for teaching me how to write.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iii
TABLE OF CONTENTS.....	v
LIST OF TABLES.....	vii
Introduction.....	1
Literature Review.....	4
Recovery.....	4
Recovery Measures.....	9
Positive Psychology.....	12
Measurement Theory.....	14
Statement of Purpose.....	26
Method.....	28
Phases of Empirical Examination.....	28
Research Questions.....	32
Results.....	33
Reliability.....	33
Single Administration Analyses.....	34
Longitudinal Analyses.....	36
Discussion.....	42
Reliability.....	43
Validity.....	44
Clinical Implications.....	51

Limitations	52
Future Directions	53
References	55
Appendix A: Utah Recovery Scale	65
Appendix B: Outcome Questionnaire (OQ [®] -45.2)	67
Appendix C: Individual item HLM analysis graphs	69
Appendix D: URS and OQ-45 HLM Treatment Trajectory Comparisons Over Time	75
Appendix E: The SAMHSA Fundamental Components of Recovery	81
Appendix F: URS Telephone Survey Script	83
Appendix G: URS Total Score Correlations with Individual OQ-45 Items	85
Appendix H: Journal Article	88

LIST OF TABLES

Table 1: Item Correlations Between the OQ-45 and the URS.....	35
Table 2: Correlations Between the quality of Life Items and the OQ-45.....	36
Table 3: Descriptive Data for the Community and Clinical Samples.....	37
Table 4: Sensitivity and Specificity of the URS Clinical Cutoff Score.....	39
Table 5: URS Change Status.....	40
Table 6: URS Total And Individual Slopes Compared to Zero.....	41
Table 7: Slope Comparisons by Change Classifications Between the URS and the OQ-45.....	42

Introduction

In 2005, the Substance Abuse and Mental Health Services Administration (SAMHSA) introduced their federal action agenda to change the system of mental health care delivery in the United States. The agenda includes several strategies for improving mental health services in the United States, such as implementing evidence-based practices, focusing on recovery, and making treatment consumer/family driven (SAMHSA, 2005). Central to SAMHSA's strategies is the idea that recovery from mental illness is not merely possible, but is rather the expected outcome of effective mental health treatment. This understanding of recovery shows how the definition of recovery has evolved (Jacobson, 2004).

Historically, recovery was thought to be a total amelioration of all psychiatric symptoms. Recovery, according to this definition, was defined by consumer's returning to a pre-morbid state of functioning. In other words, consumers were thought to have recovered when they return to a state of functioning equal to their prior functioning before the onset of any psychopathological symptoms (Davidson, O'Connell, Tondora, Lawless, & Evans, 2005). In this sense, the care provider determines recovery without consumer input.

More recently, however, a new definition of recovery has emerged which focuses on the ability of consumers of mental health services to live independently and function as contributing members in society (Jacobson, 2004). Much of the recent recovery literature has shifted towards this newer approach to recovery, which asserts that a total absence of psychopathology is unlikely to be achieved but that recovery happens when the mental health consumer is able to function normally in society through obtaining meaningful work, fulfilling relationships and independent living (Davidson, et al., 2005; Jacobson, 2004). Recovery in this sense also involves acquiring positive coping skills, restoring a sense of self, and pursuing purposeful living

(O'Connor & Delaney, 2007). It does not require a return to pre-morbid functioning; instead, mental illness is viewed as a manageable part of an otherwise complete person (Davidson et al., 2005). This definition of recovery allows for a wide range of variability as to when recovery is actually achieved. Some view it as being achieved when consumers of mental health services are able to live independently and find meaningful work, while others view it as an ongoing life-long process (Jacobson, 2004). Recovery is also considered to be an individual and subjective experience, meaning that it is difficult to define in specific terms (Jacobson & Greenley, 2001). The achievement of recovery in this sense is not determined merely by the care provider as it is in the previous definition of recovery, but rather, by a joint effort between the consumer, the care provider, and the consumer's family (Frese, Stanley, Kress, & Vogel-Scibilia, 2001; Jacobson, 2004).

SAMHSA's previously mentioned federal action agenda favors this newer approach to recovery: that recovery is a continuous process defined by independent living, meaningful work, and fulfilling relationships (SAMHSA, 2005). In order to more clearly define the construct of recovery as it pertains to the mental health community in the United States, SAMHSA organized a panel of mental health consumers, family members, providers, advocates, researchers, academicians, managed care representatives, and accreditation organization representatives which together compiled a consensus definition of recovery. The resulting consensus statement comprised 10 components that define recovery: Self-Direction, Individualized and Person-Centered, Empowerment, Holistic, Non-Linear, Strengths-Based, Peer Support, Respect, Responsibility, and Hope (SAMHSA, 2006) (see Appendix A). SAMHSA's components are purposely non-specific in order to allow for a wide range of individual recovery. They do,

however, specify a direction for mental health service providers to steer towards in terms of recovery.

In an effort to conform their mental health services to SAMHSA's 10 fundamental recovery components, the Utah division of the National Alliance on Mental Illness (NAMI-Utah) conducted focus groups of consumers in order to develop questions that could assess recovery based on both SAMHSA's model and consumer input. The focus groups operationalized the 10 components and developed 10 questions to serve as recovery indicators. The resulting 10 indicators, labeled the Utah Recovery Scale (URS), provide practitioners with a means of measuring recovery derived both from theory and consumer opinion (see Appendix A).

The recovery movement has garnered a large amount of consumer support and has begun to affect policy in the mental health treatment community. Despite the ability of this movement to effect change, very little testing as to whether the implementation of this new recovery paradigm is actually beneficial for consumers has been done. Furthermore, the research that has been done has typically relied strictly on the input of a few consumers in order to shape how recovery is measured, rather than incorporating a clear definition of the construct of recovery into their measure creation.

In addition, the existing research on recovery has also lacked theoretical and empirical grounding in terms of how the results are interpreted. Many of the components of the recovery model are not new and have been researched in other areas, such as positive psychology. In order for a recovery measure to be valid and clinically useful it must be based on a clear definition of recovery as well as consumer input. It must also be placed in a context of existing empirical research. Such a measure should be able to track changes in recovery over time so as to be beneficial for practitioners attempting to implement the recovery model.

This study is a psychometric examination of the URS which seeks to solve some of these proposed problems by comparing the URS, which was developed using both a clarified construct of recovery and consumer input, with a gold standard measure of psychotherapeutic outcome. Doing so will allow us to investigate both the psychometric utility of this new recovery instrument and the relationship between a traditional model of recovery based on symptom amelioration and this newer conceptualization of recovery. Further, this study will look to existing empirical research in the related area of positive psychology in order to provide a context in which to interpret the results of the study.

Literature Review

This section will review the construct of recovery, recovery measures, a possible empirical context for recovery—positive psychology, and measurement theory.

Recovery

Recovery has always been the central concept of psychiatric rehabilitation (O'Connor & Delaney, 2007). However, it has meant different things at different times within the mental health care community (Davidson & Roe, 2007). The first, and more traditional, meaning referred to symptom amelioration and the reduction of distress in those suffering from mental illness (Davidson, Lawless, & Leary, 2005; Davidson & Roe, 2007; N. Jacobson, 2004). In other words, recovery was thought to have occurred based on the absence of mental illness and symptomatic distress. More recently, however, a movement has arisen within the mental health community that endorses a definition of recovery based on an individual's ability to live a fulfilling and meaningful life with a mental illness (Davidson, et al., 2005; Davidson & Roe, 2007; Jacobson, 2004; Jacobson & Greenley, 2001). According to this definition, recovery is not dependent on symptom amelioration but rather an individual's ability to live a full and

meaningful life while coping with illness. This paradigm of recovery posits that individuals suffering from severe mental illness can live meaningful lives in an environment of their own choosing (O'Connor & Delaney, 2007).

According to this newer definition of recovery, an individual's acceptance of the mental illness is seen as helpful in the recovery process. Roe & Kravetz (2003) compared mental illness to a physical disability, such as paraplegia, with which an individual can learn to live a full and meaningful life. In other words, in order for a person suffering from mental illness to live a full and meaningful life they would need to make a number of adaptations to their daily living habits, much like how paraplegics would have to adapt their life style to their condition. In order for an individual to engage in this adaptation process, however, they must first accept their mental illness much as a person that has lost the use of their legs must accept their condition in order to move forward with their lives. Such an approach to recovery may not be ideal in the sense that the consumer experiences a complete absence of symptoms; however, it allows an individual to return to a somewhat normal and meaningful life (Davidson & Roe, 2007).

The fundamental aspects of recovery from this perspective are thought to relate to distress but exist on a separate continuum in the sense that an individual can grow, live meaningfully and purposefully while not living completely symptom free. In other words, mental illness is thought to be just a part of a whole person rather than the defining aspect of that individual (Davidson et al., 2005). This approach to recovery has also been described as overcoming the stigma of being a "mental patient" such as poor housing, isolation, unemployment, loss of social roles, loss of purpose in life, and iatrogenic consequences of involuntary treatment and hospitalization. It allows patients to regain some control over their lives (Davidson, et al., 2005). According to Jacobson and Curtis (2000), recovery is achieved when individuals can reclaim their own lives,

becoming autonomous individuals, competent and able to have control over their own lives. In addition, the recovery paradigm allows for consumer inclusion in services, thereby providing for a collaborative effort in facilitating recovery (Jacobson & Curtis, 2000).

This recovery model for treating mental illness has been a relatively recent development within the mental health services community and has arisen largely in response to lobbies from consumer advocacy groups such as the National Alliance for the Mentally Ill (NAMI) (Davidson et al., 2005; Davidson & Roe, 2007; Jacobson, 2004). For this reason, this new definition of recovery is thought to be more consumer-centric, as opposed to disorder centric (Frese et al., 2001). In other words, this approach to recovery focuses on the consumer as a whole, by focusing on the consumer's quality of life instead of just their symptoms. In addition, this approach is also consumer centric in that treatment is defined by the hopes and desires of the consumer as opposed to top down decisions from the health care providers. According to Frese, the recovery model emphasizes the idea that responsibility for and control of the recovery process belongs mostly to the consumer. According to Jacobson and Curtis (2000), that responsibility is not only part of the recovery process, but essential to it. This approach, which brings consumers to the table when changes to treatment are being considered, may also go a long way toward letting consumers feel that their contributions are valued and that the decision making process is fair (Frese et al., 2001).

Many researchers and practitioners have questioned the utility of the new recovery model because its components were not derived from existing theory or empirical research (Davidson, O'Connell, Tondora, Styron, & Kangas, 2006; Fisher & Ahern, 2002; Peyser, 2001; Peyser, 2001). These questions have been raised largely to call into question the ethics and economics of implementing programs based on the recovery model. Regardless of the fact that the recovery

model has arisen from consumers who feel the approaches proposed in the recovery model are missing in their treatment, there are those practitioners who do not want to spend money on new programs which are unproven in efficacy and may distract from interventions which are already shown to improve outcome.

The recovery model's political origin has made studying recovery challenging because this new approach to recovery has been operationalized and defined in a number of different ways. Further, no theoretical or empirical context has, to this point, been utilized in order to contextualize and interpret the results from recovery measures. Because of this ambiguity, recovery has been defined and measured in a myriad of different ways (e.g., in a compendium of recovery measures put together by The Evaluation Center at the Human Services Research Institute, over 42 different domains were measured in association with recovery) (Campbell-Orde, Chamberlin, Carpernter, & Leff, 2005). This wide range of domains involved in studying recovery illustrates the lack of consensus in defining recovery as a construct.

In response to the ambiguity described above, SAMHSA organized a two day conference in which 110 expert panelists consisting of mental health consumers, family members, providers, advocates, researchers, academicians, managed-care representatives, and accreditation organization representatives reviewed and discussed the recovery literature and created a 10-part consensus statement defining the fundamental components of recovery (see Appendix E) (SAMHSA, 2006). SAMHSA's resulting consensus statement on recovery provides researchers and clinicians with a more concise description of the recovery model. Further, this definition clarifies the recovery model as a construct that may lead to the development of recovery measures of greater utility for mental health service providers. SAMHSA's consensus statement on recovery organizes the recovery model into 10 fundamental components (see Appendix E).

These components, while useful, do not provide the recovery model with a bridge to existing psychological theory or empirical research. In other words, SAMHSA's definition still lacks a comprehensive theoretical and empirical base and could be thought of as a consensus statement representing a compromise among groups pushing for change. In order to address this problem, this study will utilize an empirical context of positive psychology in order to contextualize the results of the recovery measure in question. Positive psychology has been identified as a possible empirical bridge for the recovery model and provides a wide range of evidence for the types of well-being style interventions described in SAMHSA's recovery components (Resnick & Rosenheck, 2006).

In addition to an established context, the recovery model needs to be tested. Doing so will allow the model to be investigated and improved to further benefit consumers. There has been some debate, however, as to whether recovery should be measured, and what risks are involved in doing so (Jacobson & Curtis, 2000). The evidence-based model has been perceived by some in the recovery movement as stamping out hope by emphasizing an external locus of control (Frese et al., 2001). Building an evidence base for recovery, however, need not necessarily promote an external locus of control; instead, testing the model will allow it to gain more credibility and will serve to quiet many of the recovery model's critics who have complained about the models lack of empirical support. In addition, an evidence-based approach has the potential not only to provide support for the recovery model but also may serve to further refine the model to benefit consumers (O'Connor & Delaney, 2007). One of the purposes of this study is to begin to initiate such an endeavor by seeking to validate a potential measure of recovery.

According to Frese et al. (2001), the question of whether evidence-based practices can co-exist with the recovery model is an essential question to answer. They suggest that in order

for evidence-based practice to gain the support and advocacy of consumers that it is important for evidence-based practices to incorporate aspects of the recovery movement. They also suggest that researchers should be exploring ways for evidence-based practices to incorporate the subjective philosophical push of the recovery movement. One area in which the recovery model could be used is as an outcome criterion. According to Spaulding and Nolting (2006), there is broad agreement that outcome should be measured and understood in multiple domains, and that treatment of severely mentally ill populations should take multiple outcome domains that account for important aspects of real world functioning into account. One of the purposes of this study is to explore the utility of a potential measure of recovery as an outcome measure.

Recovery Measures

Researchers have attempted to measure this newer conceptualization of recovery in several ways. A literature search for recovery measures yielded three recovery measures with published reliability and validity estimates.

Recovery Assessment Scale. The Recovery Assessment Scale (RAS) is a 41-item likert type scale designed to measure recovery (Corrigan, Giffort, Rashid, Leary, & Okeke, 1999). The measure was developed from the narrative recovery stories of 4 consumers and then reviewed by an independent group of 12 consumers to ensure that the items adequately identified the construct of recovery (Corrigan, Salzer, Ralph, Sangster, & Keck, 2004). The RAS as exhibited acceptable test retest reliability ($r = .88$) and internal consistency reliability ($\alpha = .93$). In addition, the RAS has also been shown to have acceptable concurrent validity with measures of self-esteem and empowerment, as well as expected divergent validity when compared to the Brief Psychiatric Rating Scale (BPRS) (Corrigan et al., 1999). In factor analysis, five factors

emerged—personal confidence and hope, willingness to ask for help, goal and success orientation, reliance on others, and no domination by symptoms.

The RAS has established external validity in terms of recovery as individual consumers have defined it, but lacks the strength of a consensus definition of recovery. In addition, the RAS has demonstrated acceptable reliability and validity estimates but lacks any established utility for measuring changes in recovery over time.

Recovery Process Inventory. The Recovery Process Inventory (RPI) is a 22-item self-report measure that utilizes a likert-type scale to assess recovery (Jerrell, Cousins, & Roberts, 2006). The measure is based on a 10-dimensional model of recovery which was derived from four focus groups on recovery. The 10 dimensions in the model include Hope, Empowerment/Self-control, Self-Esteem, Self-management, Social Relations, Housing, Employment, Stigma, and Spirituality. The RPI has yielded internal consistency reliability estimates from .71 to .81, and test re-test reliability estimates from .36 to .63. In the factor analysis, six factors emerged: anguish, connected to others, confidence and purpose, others' care and help, good living situation, and hopeful/cares. Convergent validity with the Mental Health Statistics Improvement Program Adult Consumer Survey (MHSIP) varied from fair to moderate ($.26 \leq r \leq .55$).

While the RPI has yielded acceptable internal consistency reliability, it lacks test re-test reliability within the acceptable range for an outcome measure. In addition, the measure is only designed to be administered one time over the course of a year and has therefore not exhibited any utility for measuring change. The measure also lacks external validity in terms of recovery as a theoretical construct.

Stages of Recovery Instrument. The Stages of recovery instrument (STORI) is based on a model of recovery which incorporates four key component processes (finding and maintaining hope, the reestablishment of a positive identity, finding meaning in life, and taking responsibility for one's life) and five stages of recovery—moratorium, awareness, preparation, rebuilding, and growth) (Andresen, Caputi, & Oades, 2006). These components and stages were derived via the analysis of several personal accounts of recovery in addition to five qualitative studies that examined various stages of recovery. Concurrent validity for individual stages ranged from $r = 0.52$ ($p < 0.01$) to $r = 0.62$ ($p < 0.01$), the measure was compared to the Recovery Assessment Scale, the Mental Health Inventory, Psychological Well-Being Scales, Connor Davidson Resilience Scale, Adult State Hope Scale, and the self-identified stage of recovery as identified by the participants. The internal consistency for each of the sub-scales all returned high alpha values (from $\alpha = 0.88$ to $\alpha = 0.94$). The researchers also found evidence for divergent validity among the stages of recovery in that the most distal stages were negatively correlated and the adjacent stages were positively correlated. There was also a relationship demonstrated between the stage of recovery and a participant's score on other mental health assessments. Participants in the later stages of recovery tended to score higher on other mental health measures supporting the hypothesis that recovery is a measurable progression. While the STORI is likely useful as an ipsitive measure for identifying an individual's particular stage in recovery, it does not appear to have utility as a measure of recovery outcome in the sense that it is able to track recovery at both individual and group levels.

The recovery model is currently lacking a measure that tracks the progress of recovery in individuals and groups of individuals across time. In other words, there have been no measures that have demonstrated the utility of tracking consumer outcome based on the recovery model.

In addition, no recovery measure has been validated in a way that looks at existing theory and empirical research outside of the recovery literature. Thus it is important to place recovery in a theoretical and empirical context before explaining the methodological processes of the current study.

Positive Psychology

An area of research that could provide recovery with such a nomological context is positive psychology. According to N. Jacobson & Curtis (2000), policy makers have been looking for practical ways to incorporate the philosophy of recovery into actual practice. One way in which policy makers could go about this is by looking at existing research that may be connected to the philosophy of recovery. In addition, it is also important for empirical research to be grounded in both theory and empirical evidence. In other words, in order for effective research to be done on the concept of recovery it is important to investigate whether there is existing research on the aspects incorporated in the recovery movement to use as a context for understanding recovery research.

Like the recovery model, positive psychology theory suggests that mental health and mental illness may represent two separate but related spectrums (Keyes, 2007; Ryff, 1989; Ryff & Keyes, 1995; Seligman, Peterson, Aspinwall, & Staudinger, 2003). Resnick and Rosenheck (2006) identify a possible crossover between positive psychology and recovery. They suggest that, even though the positive psychology and recovery movements have followed separate paths, they have arrived at similar conclusions. They suggest that the recovery model could be a means for expanding the application of the positive psychology movement. In addition, Anthony (2003) argues that research on how all types of people change and grow could and should be applied to recovery.

Positive psychology research has been fueled by the idea that, in addition to abnormality, what is good in life should be studied empirically (Compton, 2005; Froh, 2004; Gable & Haidt, 2005). According to Seligman (1998), psychology should not be the study of merely what is broken but also the study what is good about life (Fowler, Seligman, & Koocher, 1999). Researchers in positive psychology argue that focusing only on the study of mental illness is an incomplete approach to psychological research because the absence of mental illness does not necessarily indicate the presence of mental health (Keyes, 2007).

Seligman first introduced the positive psychology movement as an empirical movement aimed at studying well-being in his keynote address of the 1999 annual conference of the American Psychological Association (Fowler et al., 1999). In his speech, Seligman suggested that it was not enough for psychology to focus solely on disorder: it required the study of the whole individual. In response to his speech, a movement has begun within psychology dedicated to understanding the positive side of life. The positive psychology movement has since grown exponentially, yielding hundreds of published studies and articles (Seligman, Steen, Nansook, & Peterson, 2005).

Gable & Haidt (2005) define positive psychology as a study of conditions or processes that contribute to the flourishing or optimal functioning of people. Flourishing is understood to be living within the optimal range of human functioning (Fredrickson & Losada, 2005). In other words, to flourish is to be mentally healthy. Positive psychology has focused on understanding human flourishing by focusing on such topics as strengths, virtues, resilience, well-being, and others (Gable & Haidt, 2005; Keyes, 2007; Resnick, Warmoth, & Selin, 2001; Ryff, 1989; Ryff & Keyes, 1995; Sawyer, 2002; Seligman et al., 2003).

Some researchers have called for a positive psychology approach for clinical practice, a clinical approach focusing on helping individuals to flourish rather than just on eliminating symptomatic distress (Seligman et al., 2003). These researchers suggest that it is not enough for those treating the mentally ill to focus on symptom amelioration—they must focus on helping consumers flourish. According to Keyes (2007), flourishing and mental illness can exist together in the same individual at the same time. In other words, reducing an individual's symptoms of mental illness does not ensure that there has been an improvement in their overall mental health.

Measurement Theory

In order to adequately evaluate the utility of a measure of recovery it is first necessary to conduct a cursory review of measurement theory (i.e., the rules by which psychological tests are developed). There are three categories of psychological tests: discriminative, predictive, and evaluative (Kirshner & Guyatt, 1985). Discriminative tests are used to distinguish between individuals or groups on an underlying dimension when a gold standard is unavailable (Allen & Yen, 1979). IQ tests and personality tests like the Minnesota Multiphasic Personality Inventory are examples of discriminative tests because they categorize individuals based on scores. Predictive tests are used to classify individuals into a set of predefined measurement categories when a gold standard is available (Allen & Yen, 1979). Predictive instruments can be used either concurrently or prospectively and are typically used as screening instruments to identify which specific individuals have or will develop a target condition or outcome. An evaluative test is used for the purpose of measuring the magnitude of longitudinal change in individuals or groups on a dimension of interest (Allen & Yen, 1979). Measures designed to measure treatment outcomes are evaluative because they are designed to track the progress and deterioration of individuals receiving mental health treatment.

Outcome measures have become an increasingly useful source of information about the effects of treatment for researchers, practitioners, and insurance companies in the mental health care community (Burlingame, Lambert, Reisinger, & Neff, 1995; Lambert, Gregersen, Burlingame, & Maruish, 2004). Outcome measures have been fundamental to tracking the efficacy of mental health treatment by tracking the progress of individual consumers in mental health care settings (Lambert et al., 2004; Ogles, Lambert, & Fields, 2002). They have provided researchers with a means to track the efficacy of individual therapists as well as the efficacy of different forms of treatment (Okiishi et al., 2006; Okiishi, Lambert, Nielsen, & Ogles, 2003).

Outcome measures also provide some advantages to researchers when used in place of controlled clinical trials because they allow clinicians and programs to have flexibility in how they treat individuals as opposed to requiring them to practice based on highly specific manual-based treatments (Lambert et al., 2004). This allows clinicians to approach treatment as they normally would, thus giving researchers a more accurate depiction of the effects of real treatment. Outcome measures thereby provide researchers with evidence regarding the effectiveness of treatment that is based on actual practice—practice based evidence. In addition, outcome measures provide clinicians with information regarding their clients' level of intake functioning (Lambert et al., 2001; Wells, Burlingame, Lambert, Hoag, & Hope, 1996), their progress in treatment (Lambert et al., 2001; Wells et al., 1996), and the effectiveness of specific treatment interventions (Wells et al., 1996). Consequently, outcome measures provide mental health care providers with realistic data regarding the efficacy of their programs.

In addition, increasing concern over costs and holding clinicians accountable has arisen among health care providers and corporations (Burlingame et al., 1995). Outcome measures give health care providers and third party payers with an opportunity to track therapeutic effectiveness

and make adjustments accordingly (Lambert et al., 2004; Wells et al., 1996). This allows health care providers to track the effectiveness of treatment without requiring practitioners to adhere to specific treatment models or interventions, thus allowing practitioners the flexibility to use their judgment to better help consumers. This tracking method, in turn, allows providers to track the effectiveness of treatment at multiple levels (e.g., individual, unit, and organization).

Given the potential impact of outcome measures on consumers, clinicians, and providers it is essential that such measures are both accurate and consistent in their measurements. For this reason, it is important to ensure that outcome measures yield empirical evidence for both reliability and validity (Wells et al., 1996).

Reliability. Reliability refers to how precise a test is, in other words, the degree to which test scores are consistent and repeatable (American Psychological Association., American Educational Research Association, & National Council on Measurement in Education, 1985). Reliability is defined in classical testing theory as the degree to which an observed score matches a participant's true score (Allen & Yen, 1979). In this sense, the true score is a theoretical representation of a test-taker's true response to a given test, whereas the observed score is the reported score from one administration of the test. In this sense, the reliability coefficient is one estimate of the amount of error inherent in the measure because it is an estimate of the degree to which reported scores differ from true scores (Allen & Yen, 1979). Low reliability resulting from high error variance means that the probability that the observed score reflects a participant's true score is also low. Conversely, high reliability as a result of low error variance suggests that the observed score more accurately represents the participant's true score at the time of measurement. Since a "true score" is a theoretical construct and is impossible to calculate, reliability must be estimated.

Reliability can be estimated in multiple ways. One method for estimating reliability, which focuses on examining the inter-correlations among test items, is internal-consistency reliability (Ghiselli, Campbell, & Zedeck, 1981). Internal-consistency reliability can be estimated using only one test administration and avoids the problems associated with repeated testing. The most widely used method for obtaining an internal consistency estimate for a measure that has continuous part scores is the Cronbach's Alpha statistic (Ghiselli et al., 1981). Cronbach's Alpha is a function of the ratio of the sum of the interitem covariances to the variance of the total score of the measure. This sum is largely a function of the intercorrelations among the items. Since the measure in question in the present study is thought to measure dynamic variables, internal consistency reliability is considered the appropriate method for estimating reliability (Ghiselli, et al., 1981).

Lambert, Hill, Bergin, & Garfield (1994) suggest that test-retest reliability is also particularly important to outcome research. This is because outcome measures are typically administered pre-treatment, during treatment, and post-treatment in an effort to track an individual's or group's progress in treatment. The differences in these administrations are typically calculated in order to represent some sort of change or progress in treatment. In order for a test to do this, it must demonstrate the ability to track the effects of treatment in individuals over time. Change scores are not represented solely in terms of an individual's true change score but also in terms of random measurement error (Allen & Yen, 1979). An effective outcome measure must therefore demonstrate acceptable reliability estimates in order to ensure that change scores are not merely a result of error variance. Thus, an outcome measure with acceptable test-retest reliability estimates should show change in groups receiving treatment that has been proven to be effective, and no change in groups receiving no treatment. While test-

retest reliability is important to the development of an outcome measure it is beyond the scope of the current study.

Validity. An instrument is considered valid when it can be shown to measure what it purports to measure (Allen & Yen, 1979). For example, if a measure is designed to assess a particular construct like depression, then the items should be shown to effectively measure that construct or multiple constructs if the measure is designed to measure multiple constructs. There are a number of ways in which validity estimates can be calculated, including construct validity, content validity, and criterion-related validity (Allen & Yen, 1979).

Construct validity is the degree to which an instrument accurately measures the theoretical construct that it is designed to measure. A combination of various validity estimates makes up the construct validity of an instrument. According to Allen and Yen (1979), construct validity is an ongoing process and may include any stable prediction, both criterion and content related.

Content validity is determined by subjective judgment based on rational analysis of the content of a test. There are two main types of content validity: face validity, in which the test items appear to measure the construct they purport to measure, and logical validity, in which a panel of experts examines a measure according to specific criteria (Allen & Yen, 1979). Content validity is not considered sufficient justification for a test's use (Allen & Yen, 1979).

Criterion-related validity is determined by how well a test score predicts some type of criterion or human behavior. In the case of recovery the criterion we are using is traditional outcome. Theoretically, a high score on a recovery measure should correlate moderately with symptom amelioration. There are two types of criterion-related validity: predictive validity, which looks at how well a test predicts some type of future behavior, and concurrent validity,

which looks at how well a test predicts behavior that is co-occurring at the time the measure is given. Predictive validity involves using test scores to predict future behavior. A predictive validity coefficient is obtained by giving the test to all relevant people, waiting a reasonable amount of time, collecting criterion scores, and calculating the validity coefficient (Allen & Yen, 1979). This study will seek to explore the construct validity of a recovery measure in the areas of criterion-related validity in the form of concurrent validity. Concurrent validity is obtained by calculating the validity coefficient of two measures given at the same time.

Since construct validity is the degree to which a measure actually measures the trait or characteristic that it purports to measure, it is important to assess the degree to which an outcome instrument accurately identifies individuals that fall into clinical and non-clinical ranges (i.e., identify individuals that are more likely to be in need of mental health services). If a measure is unable to distinguish between individuals that are likely to fall in either a clinical or non-clinical population, then it would also not be useful in terms of distinguishing when an individual has moved from being likely to fall in a clinical population vs. falling in a non-clinical population. One way to investigate how well a measure does this is by calculating a clinical cut-off point. The cutoff point is the numerical score between adjacent samples where it is statistically more likely for a score to be in one, as opposed to the other adjacent overlapping sample (Jacobson et al., 1984). A clinical cutoff score can be calculated by utilizing the means and standard deviations of the two differing populations (in this case individuals in a clinical population vs. individuals that are not). The equation for calculating the cutoff score is as follows:

$$c = \frac{(SD_1)(Mean_2) + (SD_2)(Mean_1)}{SD_1 + SD_2}$$

Where SD_1 is the standard deviation of the clinical (consumer) population, $Mean_2$ is the mean of

the non-clinical population, SD_2 is the standard deviation of the non-clinical population, and $Mean_1$ is the mean of the clinical population. Once a cutoff score has been calculated it is important to assess its accuracy in terms of its ability to identify individuals that fall in each population.

One of the ways to do this is by calculating the sensitivity and specificity of the instrument. Both sensitivity and specificity were defined first by Yerushalmy (1947) and have been an important part of the medical literature since that time (Griner, Mayewski, Mushlin, & Greenlan, 1981; Vecchio, 1966). Sensitivity, in this sense, can be understood as the degree to which a measure accurately identifies individuals that possess the attribute of interest. In this case, it would be the degree that a recovery measure accurately identifies consumers who function in the clinical range, obtaining scores in the clinical range of the screening test as determined by the cutoff score. Sensitivity is calculated by dividing the number of true positives (individuals accurately identified as falling within the clinical range) by the sum of true positives and false negatives (individuals inaccurately identified as not falling within the clinical range). Specificity refers to the degree to which a measure is able to accurately identify consumers that fall in the non-clinical range. It is calculated by dividing the number of true negatives by the number of true negatives added to the number of false positives.

Sensitivity and specificity are estimated using different samples of people and vary independently of one another. For this reason both sensitivity and specificity can be high and approach the perfect case of a test that is 100% sensitive and 100% specific, or both can be low. In addition, both indices are independent of sample size and population base rates.

In addition to sensitivity and specificity, it can also be helpful to examine the positive predictive power and negative predictive power of an instrument (Griner et al., 1981). The

positive predictive power of a test is the likelihood that a person with a positive test finding actually falls into the identified population, in this case whether someone identified as belonging in the clinical population actually is in a clinical population. Positive predictive power is calculated by dividing the number of actual positives by the number of total positives, including false positives, identified by the test. The negative predictive power of a test is the likelihood that an individual that is not identified as falling within the clinical population actually does not fall into the clinical population. Negative predictive power is calculated by dividing the actual number of negatives by the total number of negatives, including false negatives, identified by the test. When looking at both positive predictive power and negative predictive power, it is important to consider that neither estimate is independent of sample size: both high and low calculations of positive and negative predictive power could be the result of the number of individuals that fall in either the clinical or non-clinical ranges.

Validity for change. Outcome measures are designed to track change in individuals resulting from treatment. A measure may be a valid indicator of a characteristic without being able to measure change in that characteristic (Lambert et al., 1994). For this reason, in order for an outcome measure to be valid, its ability to detect change be demonstrated. According to Lipsey (1983), validity for change can be demonstrated from the effect size when comparing treatment groups and comparison groups from the same populations. The effect size for an outcome measure is determined by calculating the difference between post-treatment means for treatment and comparison groups divided by the outcome measure's standard deviation. In other words, the effect size as represented by d is a measure of the degree to which population means of two samples differ ($\mu_1 - \mu_2$). Effect sizes are considered small when $d = .20$, moderate when $d = .50$, and large when $d = .80$ or greater.

Sensitivity to change. Sensitivity to change, or responsiveness, was first emphasized as an important psychometric property in the 1970's (Aiken, 1977). Guyatt, Walter, and Norman (1987) defined responsiveness as an instrument's ability to detect clinically important change and suggested that a measure's sensitivity to change is determined by two properties. First, sensitive measures must yield more or less the same scores when subjects are stable. Second, it must register score changes when subjects' health status improves or deteriorates. They further suggested that failure to demonstrate responsiveness is the product of one of two factors: either treatment did not work or the instrument used was inadequate in assessing changes that occurred. Deyo, Diehr, and Patrick (1991) viewed responsiveness as the ability of an instrument to detect small but important clinical changes. Kazdin (1998) defined sensitivity as the ability of a dependent measure to be sensitive to the type and magnitude of change that the investigator is expecting. In this study, we expect to see a correlation between recovery and traditional outcome. In other words, when consumer scores remain stable on a gold standard measure of outcome, we would expect to see their recovery scores remain stable as well. In addition, as consumers show improvement or deterioration as measured according to a gold standard outcome measure, we would expect to see that improvement or deterioration reflected in their scores as measured by the recovery measure.

In addition, an outcome measure must be sensitive to clinically significant change (Jacobson, Truax, & Kazdin, 1992). Clinically significant change has been defined as a post-treatment return to normal functioning (Jacobson & Truax, 1991). This type of change can be evaluated in three ways (Jacobson et al., 1992). First, post-treatment functioning of consumers should fall outside the range of the dysfunctional population with the range defined as two standard deviations beyond the mean for that population in the direction of functioning. Second,

post-treatment consumer functioning should fall within the range of the functional population with the range defined as within two standard deviations of the mean of that population. Third, post-treatment consumer functioning should be closer to the mean of the functioning population than the mean of the dysfunctional population.

This conceptualization of clinically significant change is somewhat problematic since the recovery model posits that a return to normal functioning is unnecessary in order for consumers to meet the criteria of having recovered. For this reason, there is a strong possibility that the distributions of functional and dysfunctional as defined by the recovery model will exhibit a great deal of overlap. Thus, it is essential to ensure that the changes measured by an instrument are reliable (Jacobson et al., 1992). In order to ensure that this is the case, a reliable change index (RCI) can be calculated. An RCI is calculated by subtracting consumer initial treatment scores from post-treatment scores and then dividing them by the standard error of the difference. The standard error of the difference describes the spread of the distribution of change scores that would be expected if no actual change had occurred and is calculated by taking the square root of twice the standard error of measurement of a distribution squared. According to Jacobson et al. (1992), an RCI greater than 1.96 would be unlikely to occur ($p < .05$) without change that was not due to chance alone. If the RCI is less than 1.96 then the measured change may be due to the fluctuations of an imprecise measurement. Calculating an RCI will allow us to examine whether the recovery measure in question is able to detect clinically significant changes in consumers that are greater than would have occurred by chance alone (Jacobson et al., 1992).

Analysis of therapeutic change. Statistical methods are particularly important when assessing change because different methods of analysis on the same data can lead researchers to come to different conclusions about client data (Speer & Greenbaum, 1995). In order to evaluate

change, a statistical procedure must be able to accommodate a multiple administrations or repeated measures design. In addition, the method must be able to detect change at both the individual and group levels.

Traditionally, univariate (ANOVA) or multivariate (MANOVA) analysis of variance procedures have been used to track outcome (Raudenbush & Chan, 1993). According to Ware (1985), the aforementioned analyses are inappropriate when studies of change contain missing data, time-varying covariates, unbalanced designs or continuous predictors of rates of change. In addition, these models do not directly model for individual variation, but instead account for it only within interactions of repetitions. In contrast, hierarchical linear modeling (HLM) offers an analytic approach without the limitations as MANOVA and ANOVA (Raudenbush & Chan, 1993).

HLM estimates linear equations that explain outcomes for members of groups at both individual and group levels (Arnold, 1992). The models are hierarchical because they involve predicting the characteristics of members who are nested within a group or a network of groups. Each group may then be encompassed within a larger group. At each level, each member of the group belongs to only one group at the next hierarchical level (Arnold, 1992). Most mental health services are provided within nested groups, and a great deal of research on mental health care involves tracking the effect that specific groups have on outcome. Consumers are often placed in the care of therapists within mental health care systems that are often part of larger health care organizations, and factors at each level of care delivery can affect consumer outcomes. Since consumer growth or deterioration may occur within all of the aforementioned levels, identifying the predictors of consumer growth is a multilevel problem.

HLM involves performing regressions of regressions (Arnold, 1992). Therefore, it presumes a working knowledge of linear regression analysis and its assumptions. The regressions are done at the lowest unit of analysis (for example, consumers) within the next higher unit of analysis (for example, therapists) and so on. At the first level, regression equations for each consumer predict consumer outcomes as a function of other consumer characteristics within each therapist or treatment center. These equations are called “within-unit” models, and there is one for each treatment center. The intercepts and coefficients in these equations usually vary randomly across individuals and are used as the dependent variables in second-level regression equations with individuals as the unit of analysis and individual characteristics as the independent variables. These regression equations are called “between-unit” models. HLM differs specifically from other slopes-as-outcomes methods in that HLM accounts for the variance around each parameter from the first level in the regression analyses at the next level (Arnold, 1992).

Research examining treatment trajectory growth curves is ideally suited to the use of HLM that can be generated from consumer responses to outcome measure items (Arnold, 1992). HLM growth curves allow the development of models of individual growth as well as the study of differences between individuals in that growth because it allows for the examination of the data at multiple levels. In addition, utilizing HLM for examining treatment trajectories in this way has several advantages for use in outcome research over regression or MANOVA. HLM’s use of the expectation-maximization algorithm accounts for missing data so that missing participant data does not need to be thrown out because of the limitations of the model (Speer & Greenbaum, 1995).

In addition, HLM exhibits greater precision for assessing individual and group changes because of the use of Bayesian estimation (Raudenbush & Chan, 1993; Speer & Greenbaum, 1995). HLM also is more flexible in terms of its data requirements because of nesting (Raudenbush & Chan, 1993; Speer & Greenbaum, 1995). The repeated observations in HLM are hierarchical, and thus participants may be analyzed at different times and on varied occasions—each observation being viewed as nested within each individual participant (Bryk & Raudenbush, 1987). In other words, HLM allows for the examination of a longitudinal data set that utilizes consumer data comprised of varying waves of data points and administration intervals. HLM is therefore an improvement on ANOVA and MANOVA because it utilizes all available data (Speer & Greenbaum, 1995).

HLM does, however, require that certain assumptions about the data be satisfied for its use. The first level of analysis requires at least three within-subject data points. The data must consist of units nested within groups (hierarchical). The data from the first level of analysis form the basis for the second level of analysis so they must be highly reliable and valid. The groups must have enough within-subjects and between-subjects classifications to provide adequate degrees of freedom. Large samples are recommended, but specifications on how large are not given in the literature (Arnold, 1992). HLM involves a regression of a regression therefore no assumptions about causation can be applied to HLM results (Arnold, 1992).

Statement of Purpose

The present study seeks to investigate the relationship between the new and old conceptualizations of recovery and whether consumers with aspects of the recovery model emphasized in their treatment actually experience greater symptom amelioration than consumers not receiving that type of treatment. This study will be correlational, so it will not be able to

establish a controlled treatment comparison; however, it will provide an initial empirical opportunity for investigating the recovery model. In order to assess whether consumers are progressing in treatment according to the standard outlined in the recovery model, it is necessary that a measure yield both acceptable reliability and validity estimates. In addition, it is necessary that such an instrument demonstrate the ability to detect changes in consumers over time. The purpose of this study is to address the reliability and validity of the Utah Recovery Scale (URS). Because no existing quantitative research has been done utilizing the URS, the aims of the study can be better understood in terms of research questions rather than hypotheses:

1. Is there evidence supporting the reliability of the URS?
 - a. What evidence is there for the internal-consistency reliability of the URS (.8 or greater)?
2. Is there evidence supporting the construct validity of the URS as a measure of mental health treatment outcomes?
 - a. What evidence is there for the concurrent validity of the URS?
 - b. What is the relationship between recovery as measured by the URS and traditional mental health treatment outcome as measured by the OQ-45?
 - c. What is the relationship between recovery as measured by the URS and well-being as measured by the quality of life items on the OQ-45?
 - d. Is there evidence that the URS can distinguish between consumers that fall in clinical and non-clinical ranges?
 - e. What is the sensitivity and specificity of the URS?
 - f. Is there evidence that the URS is sensitive to clinically meaningful changes in consumers?

- g. Are changes in individual consumers over time using the URS and its individual items significantly different from zero?
- h. Are the changes in consumers observed utilizing the URS greater than would have occurred by chance alone as indicated by the reliable change index of the instrument?
- i. How do changes observed in consumers using the URS compare to changes in the same consumers using the OQ-45?

Method

Phases of Empirical Examination

In order to answer the above research questions it was necessary to examine the URS in two phases. The first phase of the study was to collect community data for the URS in order to obtain URS data of a non-clinical sample. The next phase of the study was to examine archival URS and OQ-45 data obtained through the state of Utah.

Phase 1: Community Sample. In order to calculate a clinical cut point, URS data from a non-clinical community sample was collected.

Participant selection. Participants were selected randomly using a local phone book. Trained researchers contacted participants via telephone and administered the survey according to an administration script (see Appendix F). A total of 91 participants were sampled including 53 women and 38 men. Participant ages ranged from 18 to 86 with the average age being 44. Data was not collected for individuals under the age of 18. Demographically, 84 participants identified themselves as Caucasian, 2 as African American, 2 as Asian, 1 as Polynesian, 1 as Latino, and 1 as Mixed Race. All participants indicated that they were not currently receiving any form of mental health treatment.

Instrument. The initial phase of the study examined the data collected using the Utah Recovery Scale (URS). The URS was derived from a series of consumer focus groups conducted by the Utah division of the National Alliance for the Mentally Ill (NAMI-Utah) in order to construct questions based on SAMHSA's fundamental components of recovery. The focus groups resulted in 10 questions based on SAMHSA's fundamental recovery components of Self Direction, Individualized and Person-Centered, Empowerment, Holistic, Non-Linear, Strengths Based, Peer Support, Respect, Responsibility, and Hope (SAMHSA, 2005). The URS is scored using a 5-point scale (0= never 1 = rarely, 2 = sometimes, 3= frequently, 4 = almost always), which yields a possible range of scores from 0 to 50.

Data collection. The URS currently has no developed reliability or validity statistics. Initial criterion validity for the measure, however, was investigated using additional consumer focus groups which pilot tested the measure. Participants for three focus groups were selected at three different community mental health locations. Two of the focus groups consisted of patients receiving inpatient treatment; the other consisted of participants receiving outpatient treatment. Focus groups ranged from 5 to 15 participants, comprising of a total of 30 participants. The focus group participants, after granting their consent to participate in the focus groups, were asked to take the URS and then were asked questions about their experience taking the survey, the survey questions, and about what recovery means to them in general. The focus group leader closely followed a discussion guide of questions and asked follow-up questions when appropriate. The groups lasted between 35 and 50 minutes and were digitally recorded for further analysis. The discussion leader reviewed the digital recordings in conjunction with his notes from the focus groups and analyzed them for emerging themes that were common among focus groups.

Response to the URS from the focus groups was, for the most part, overwhelmingly positive. Almost all of the consumers felt that the questions addressed things that were important to them—something that they did not always feel other outcome measures they had been exposed to had done. Many also expressed that the measure addressed specific concerns that were important to them and to their treatment.

Phase 2: Clinical sample. In order to calculate a clinical cut point, RCI, sensitivity and specificity, and longitudinal change trajectories a clinical archival data sample was also examined.

Archival data. Consumer data for this study was taken from the archival records of community mental health centers in the state of Utah. Utah regularly tracks consumer outcome at community mental health centers that receive state funding using a variety of measurement tools including the Outcome Questionnaire-45 (OQ-45) and the URS. Consumer data are then databased and reported back to clinicians using the OQ-45 analysis system that was developed and supported by OQ-45 Measures LLC. This database provides patient-level demographic data combined with outcome data for each consumer. Approximately 30,000 adults receive mental health services from the state services on a yearly basis. Outcome data from inpatient, residential, and outpatient consumers was analyzed.

Instruments. In addition to the URS, this phase of the study also examined consumer distress as measured by OQ-45, which is designed to measure client outcomes in a therapeutic setting (Lambert et al., 2004). The OQ-45 is a 45-item self-report questionnaire scored using a 5-point scale (0= never 1 = rarely, 2 = sometimes, 3= frequently, 4 = almost always) and yields a possible range of scores from 0 to 180. High scores on the OQ-45 indicate more distress and as clients improve scores decrease. The OQ-45 has three subscales that measure the quality of

interpersonal relations, social role functioning and symptomatic distress. The OQ-45 has been validated across cultures using a variety of normal and consumer populations (Lambert et al., 2004). The OQ-45 has become a gold standard for measuring symptomatic distress as it relates to treatment, and was recently found to be the third most commonly used measure of outcome in a survey by psychologists in clinical practice (Hatfield & Ogles, 2004). In addition, the OQ-45 has demonstrated utility for tracking a client's progress in therapy as well as measuring therapist effectiveness (Okiishi et al., 2006; Okiishi et al., 2003; Vermeersch, 1998; Vermeersch, Lambert, & Burlingame, 2000; Vermeersch et al., 2004).

The OQ-45 takes approximately five to seven minutes to complete and is typically administered prior to each treatment session. Concurrent validity is moderate to high ($r = 0.50-0.85$) when correlated with measures most often used to assess psychotherapy outcome in clinical trials (Lambert et al., 2004). Most importantly, the OQ-45 has been shown to be sensitive to changes in clients over short time periods while remaining stable in untreated individuals (Vermeersch, Lambert, & Burlingame, 2000; Vermeersch et al., 2004). The OQ-45 has a reported three-week test-retest reliability value of $r = .84$ and a reported internal consistency of $r = .93$ (Lambert et al., 2004).

Reliable change indices (RCI) for the OQ-45 have been calculated using formulas developed by N. S. Jacobson and Truax (1991). The RCI for the OQ-45 was calculated to be 14 points using normative data from the community non-clients ($N = 1353$) and clients entering treatment ($n = 1476$), thus clients who exhibit a 14 point positive or negative change are found to have made reliable change. A clinical cutoff score on the OQ-45 was found to be 63; thus, when a client's score drops below 63, they are thought to be functioning more like non-clients than typical client populations. When a client's score has dropped by 14 points or more and the

clinical cutoff score has fallen below 63, the client has met the criteria for clinically significant change (Ellsworth et al., 2006).

In addition to the total score, the OQ-45 has three subscales that measure quality of interpersonal relations, social role functioning and, symptom distress (Lambert, et al., 2004). The interpersonal relations subscale consists of 11 items that attempt to assess functioning in interpersonal relationships. The social role subscale consists of nine items that measure dysfunction in roles such as work and leisure life activities. The final subscale, the symptom distress subscale, is a 25-item scale that evaluates symptoms such as depression and anxiety.

Research Questions

In order to assess whether consumers are progressing in treatment according to the standard outlined in the recovery model, it is necessary that a measure yield both acceptable reliability and validity estimates. It is also necessary that such an instrument demonstrate the ability to detect changes in consumers over time. The purpose of this study was to address the reliability and validity of the URS in addition to the measure's ability to detect clinically meaningful changes over time. Since no prior empirical data has been collected for the URS, this study is better understood in terms of the following research questions, rather than directional hypotheses, were explored:

- What is the internal consistency reliability for the URS?
- What is the relationship between the URS and traditional outcome as measured by a gold standard measure of outcome (OQ-45)?
- Is the URS capable of measuring clinically meaningful change as statistically defined using a Reliable Change Index (Jacobson & Truax, 1991).

- How do meaningful changes in consumers as measured by the OQ-45 over a six-month period and changes measured in the same consumers over the course of the same time period by the URS relate to one another?

Several analyses were used in order to answer these questions. First, the internal consistency of the URS was calculated from the non-clinical data sample using the Cronbach's Alpha statistic. Second, correlational procedures were used to compare the URS and the OQ-45 based on a one-time administration of the measures in a clinical consumer sample. Third, both the clinical and non-clinical samples were used to calculate a cutoff score. Fourth, clinical longitudinal data were used to calculate an RCI for the URS that was then used to compare consumer change classifications on both instruments. Lastly, consumer change trajectories on both measures were examined and compared utilizing hierarchical linear modeling.

Results

Data for this study was examined in three areas: reliability analyses from a community sample, single administration analyses exploring the relationship between the URS and the OQ-45 in a clinical sample, and longitudinal analyses of the relationship between consumer change as measured by the two instruments.

Reliability

In order to explore the reliability of the URS, participants were randomly selected from the community. After agreeing to participate in the study and giving consent participants answered a few demographic questions and then the 10 URS items over the phone. The internal consistency for the URS scores was estimated using the Cronbach's alpha statistic. Six participants who indicated that they were currently receiving some form of mental health

treatment were excluded from this analysis. The overall URS internal consistency estimate was high ($\alpha = 0.898$, $n = 91$).

Single Administration Analyses

Archival single administration consumer data from community mental health centers in Utah were used to explore the relationship between the URS and the OQ-45. This data was explored in terms of an overall relationship between the two measures, the relationship between the subscales of the OQ-45 and the URS, and the relationship between the quality of life items on the OQ-45 and the URS.

Initial analyses indicated a strong inverse relationship between the two measures ($r = -0.75$, $n = 8,483$). This indicates that consumers who scored higher on recovery as measured by the URS scored lower on distress as measured by the OQ-45. In addition, several individual items from the two measures had moderately strong negative correlations ($r < -.6$), which are reported in Table 1.

In addition, moderate to strong relationships were also found between the URS and each of the OQ-45 subscales: Symptom Distress ($-.711$, $n = 7878$), Social Role Functioning ($-.576$, $n = 8,197$), and Interpersonal Relations ($-.746$, $n = 8162$). Consumers with missing scores on these particular items were excluded from this analysis.

The OQ-45 also contains several positively worded items that are reverse scored that can be considered quality of life items. As with the rest of the OQ-45, higher scores on these items are indicative of higher levels of distress. A composite score was created from these items and then compared with the URS total score. A strong inverse relationship was shown between the quality of life items (higher scores indicate higher levels of distress) and the URS (higher scores indicate higher levels of recovery; $r = -.863$, $n = 8,216$). Consumers with missing scores on these

Table 1

Item Correlations Between the OQ-45 and the URS

URS items	OQ-45 items	r	n
1. I have hope for the future.	13. I am a happy person.	-0.634	8739
	15. I feel worthless.	-0.615	8738
	23. I feel hopeless about the future.	-0.671	8746
	24. I like myself.	-0.643	8742
	31. I am satisfied with my life.	-0.67	8761
	OQ-45 total	-0.67	8750
2. Most days I get to do something that I enjoy.	—	—	—
3. I do something I enjoy during the day.	21. I enjoy my spare time.	-0.65	8726
	31. I am satisfied with my life.	-0.611	8744
	OQ-45 total	-0.619	8734
	—	—	—
4. I feel the place I live is ok.	—	—	—
5. My life has meaning.	13. I am a happy person.	-0.624	8714
	20. I feel loved and wanted	-0.617	8731
	24. I like myself.	-0.638	8716
	31. I am satisfied with my life.	-0.683	8736
	OQ-45 total	-0.648	8724
6. I have people/friends that I can turn to.	20. I feel loved and wanted	-0.612	8722
7. I am connected to my community.	—	—	—
8. I am in charge of my own life and recovery.	—	—	—
9. I have goals for my future.	—	—	—
10. My relationships are meaningful.	20. I feel loved and wanted	-0.632	8717
	43. I am satisfied with my relationships with others.	-0.617	8705
	3. I feel no interest in things.	-0.612	8474
URS total	13. I am a happy person.	-0.709	8483
	20. I feel loved and wanted	-0.702	8491
	21. I enjoy my spare time.	-0.663	8483
	24. I like myself.	-0.706	8487
	42. I feel blue	-0.608	8495
	43. I am satisfied with my relationships with others.	-0.692	8486

Note. — indicates that correlations for that item were weaker than .6.

Table 2

Correlations Between the Quality of Life Items and the OQ-45.

URS	OQ-45	r	n
1. I have hope for the future.	13 I am a happy person.	-0.634	8739
	24 I like myself.	-0.643	8742
	31 I am satisfied with my life.	-0.67	8761
3. I do something I enjoy during the day.	21 I enjoy my spare time.	-0.65	8726
	31. I am satisfied with my life.	-0.611	8744
5. I have meaning in my life.	13 I am a happy person.	-0.624	8714
	24 I like myself.	-0.638	8716
	31 I am satisfied with my life.	-0.663	8736
10. My relationships are meaningful.	20 I feel loved and wanted.	-0.632	8717
	43 I am satisfied with my relationships with others.	-0.617	8705
URS Total	Positive item composite score	-0.863	8216

particular items were excluded from this analysis. Individual items that were found to have a moderately strong relationship ($r < -.6$) are reported in Table 2.

Longitudinal Analyses

Archival longitudinal consumer data from community mental health centers in the state of Utah were explored in order to assess the URS's ability to detect clinically significant changes among consumers. The data was examined in several ways including the calculation of a cutoff score, a reliable change index, and calculations of longitudinal consumer change trajectories. The longitudinal sample used for these analyses consisted of 1,445 consumers that took the URS over several administrations ranging from 2 to 16, with the average number of administrations being 3.5.

Cutoff score. The first step in this process was to calculate a cutoff point—the point along the range of possible URS scores that best separates the distribution of the non-consumer

population scores from the distribution of consumer population scores. The cutoff point between adjacent samples defines the point where it is statistically more likely for a score to be in one, as opposed to the adjacent overlapping distribution (Jacobson et al., 1984). The calculation is as follows:

$$c = \frac{(SD_1)(Mean_2) + (SD_2)(Mean_1)}{SD_1 + SD_2}$$

Where SD_1 is the standard deviation of the clinical (consumer) population, $Mean_2$ is the mean of the non-patient population, SD_2 is the standard deviation of the non-patient population, and $Mean_1$ is the mean of the clinical (consumer) population. In order ensure homogeneity in the consumer sample, consumers whose OQ-45 score fell below the clinical cutoff of 63 were excluded from the analysis. Descriptive data for these two samples on the URS are reported in Table 3.

Table 3

<i>Descriptive Data for the Community and Clinical Samples</i>			
Sample	Mean	SD	n
Clinical	19.67	7.072	5,010
Community	35.84	5.455	91

Prior to calculating a cutoff point, it was necessary to ensure that both samples used in the calculation were statistically distinct. Distinctness of the samples was determined using an independent samples t test and a “ d ” test. The two samples met the two criteria for statistical distinctness. The independent samples t value of -27.841 ($p < .05$, $df = 95.577$) surpassed alpha of .05. Equal variances were not assumed for this calculation providing a more conservative estimate of t . In addition, the calculated “ d ” value surpassed the criterion of .5, further indicating the distinctiveness of the two samples ($d = 2.56$). Once distinct samples were identified and

statistically verified, the next step of generating a cutoff point was completed. Using the above formula a cutoff score of 28.799 (28/29) was calculated.

The clinical cutoff score enables the URS to distinguish consumers from non-consumers. Table 4 presents the accuracy with which the cutoff score classified respondents. The sensitivity of .89 indicates that 89% of actual patients scored below the clinical cutoff score of 29 and 11% scored above. In other words, use of the cutoff score identified 89% of actual consumers. The specificity of .95 indicates that of the actual non-consumers, 95% scored above the cutoff score and 5% scored below. Thus the cutoff score identified 95% of actual non-consumers. The overall accuracy of the cutoff score in predicting patient versus non-patient status was 0.89 (hit rate). The positive predictive power (PPP) of .99 indicates that of the cases the cutoff score predicted to be patients, 99% of them were actual patients. The negative predictive power (NPP) of .14 indicates that of the cases the cutoff score predicted to be non-patients, 14% of them were actual non-patients. Note that whereas the test characteristics of sensitivity, specificity, and hit rate are relatively independent of prevalence of the condition being tested, the characteristics of PPP and NPP are highly dependent upon prevalence (Streiner, 2003). Thus the high PPP and low NPP are expected considering the large proportion of consumers (98%) versus non-consumers (2%) in the sample producing these indices.

Table 4

		Actual		Total
		Positive (consumer)	Negative (non-consumer)	
Predicted	Positive (consumer)	True Positives 86	False Positives 536	622 12%
	Negative (non-consumer)	False Negatives 5	True Negatives 4474	4479 88%
Total		91 2%	5010 98%	5101 100%

Sensitivity	Specificity	Hit Rate	Positive Predictive Power	Negative Predictive Power
0.89	0.95	0.89	0.99	0.14

Reliable change index. In addition to a cutoff score, a reliable change index (RCI) was also calculated for the URS. The RCI is a way of determining whether changes observed on a measure are greater than what would be attributed to measurement error. The RCI is calculated by multiplying the standard error of the difference by the point location on a distribution in order to achieve a certain confidence interval. The RCI for the URS was calculated at the 95% confidence level. In addition, the reliability coefficient used in the equation was Cronbach's alpha coefficient of .898, obtained from the community sample. The equation for calculating the RCI is presented below:

$$95\%RCI = S_{diff} = (4.095)(1.96) = 8.026 = 8/9$$

$$S_{diff} = \sqrt{2(S_E)^2} = \sqrt{2(2.896)^2} = 4.095$$

$$S_E = SD_p \sqrt{1 - \alpha} = 9.066 \sqrt{1 - 0.898} = 2.896$$

$$SD_p = \sqrt{\frac{(n_1 - 1)(SD_1^2) + (n_2 - 1)(SD_2^2)}{(n_1 + n_2 - 2)}} = \sqrt{\frac{(1542 - 1)(8.983) + (1445 - 1)(9.1554)}{(1542 + 1445 - 2)}} = 9.066$$

where 95% RCI is the RCI value at the 95% confidence level, S_{diff} is the standard error of the difference, S_E is the standard error of measurement, SD_p is the pooled standard deviation for the first and last score used in the calculation, α is the internal consistency reliability utilized for the calculation, n_1 is the sample size for the first administration, n_2 is the sample size for the last administration, SD_1 is the standard deviation of the first administration of the URS, and SD_2 is the Standard deviation of the last administration of the URS. The calculated RCI for the URS was 8.026 indicating that consumers who have changes greater than 8 points in their total URS score have made clinically significant change.

The RCI and the cutoff score were then used in order to determine consumer improvement, deterioration, and stability. In addition, consumers that reached the clinical cutoff score of greater than 31 were considered recovered. Consumers whose first and last scores did not have a corresponding OQ-45 score because of missing data were excluded from this analysis. The results are presented in table 5.

Table 5

<i>URS Change Status</i>		
	n	% of total
Total	1445	100%
Recovered	176	12%
No change	1114	77%
Improved	67	5%
Deteriorated	88	6%

Change trajectories. Longitudinal consumer change trajectories were also examined. First, consumer URS change trajectories were compared with zero (no change), and lastly consumer URS change trajectories were compared with consumer change trajectories on the OQ-45.

In order to examine whether changes observed on the URS were significantly different from no change consumer URS change trajectories were compared to zero or no change. HLM was used for this analysis in order to account for the multiple waves of consumer data as well as missing data. All but three (items 4, 6, and 10) of the URS items were shown demonstrate change that was significantly different from zero. The results are presented in Table 6.

Consumer change trajectories on the URS were then compared to consumer change trajectories on the OQ-45 in order to explore whether the two measures appear to be measuring change in the same way.

Table 6

<i>URS Total and Individual Item Slopes Compared to Zero</i>		
Item	Slope	t
Total	0.07654	5.87*
1. I have hope for the future.	0.04967	3.18*
2. I have meaningful work/volunteer activities in my life.	0.06763	4.03*
3. Most days I get to do something that I enjoy.	0.07075	4.47*
4. I have a place to live and it's ok.	0.01913	1.16
5. My life has meaning.	0.0747	5.16*
6. I have people/friends I can turn to.	0.02135	1.36
7. I am connected to my community	0.075	4.99*
8. I am in charge of my own life and recovery.	0.07862	4.57*
9. I have goals for the future.	0.05082	3.48*
10. I have meaningful relationships	0.02817	0.078

$N = 1954$

* $p < .05$.

Prior to conducting these statistical analyses, reverse scoring procedures were performed so that increasing scores corresponded to increasing levels of psychopathology on all URS items.

In addition, scores from the two tests were standardized using Z scores so that both measures were on an equal scale. Again, HLM was used for the analyses because of the multiple waves of consumer data and to account for missing data. Results of the initial data analysis indicated that in the overall consumer sample the URS seems to track change in the theoretically proposed

direction (i.e., consumers improved over time at a similar rate as measured by the OQ-45). In groups in which the slopes of the two measures were statistically different, the OQ-45 consistently had a steeper slope. The results of these comparisons are presented in Table 7 and Appendix D.

Table 7

Slope Comparisons by Change Classification Between the URS and the OQ-45

Classification	Slope URS	Slope OQ-45	n	t
Total	-0.079204	-0.08045	1954	0.09
Distressed	-0.14791	-0.1921	1151	2.74*
Distressed improved	-0.3765	-0.5247	203	4.07*
Distressed deteriorated	0.37348	0.4402	142	-1.23
Distressed no change	-0.04298	-0.02063	612	-1.19
Distressed recovered	-0.5831	-0.7751	190	4.19*
Not distressed	0.03121	0.09269	798	-2.48*
Not distressed improved	-0.2646	-0.4318	122	2.76*
Not distressed deteriorated	0.7459	0.4733	177	-4.65*
Not distressed no change	-0.03566	0.01192	498	-1.73
Beginning treatment	-0.12559	-.016657	685	1.53

Note. 5 cases could not be classified as distressed or not distressed because they were missing their initial OQ-45 score and were therefore excluded from subsequent analyses.

5 additional consumers could not have change statuses calculated because of missing data and were excluded from subsequent analyses.

* $p < .05$.

Discussion

The purpose of this study was to psychometrically examine the utility of the URS as an evaluative measure for mental health treatment services. Evaluative measures such as the URS, which are designed to track treatment progress, have the potential to greatly impact the treatment that consumers receive by providing incremental feedback to both clinicians and service administrators regarding client progress in treatment. Given this potential impact of evaluative measures on consumers, clinicians, and providers, it was necessary to explore both the reliability and the validity of the URS as an evaluative measure. The URS was examined in several

psychometric categories including reliability, validity, sensitivity to change, and validity for change. Two sets of URS data were utilized in order to examine questions of reliability and validity about the instrument. First, data were collected from a randomly sampled community population that was not receiving any kind of mental health treatment in order to explore the reliability of the URS. Next, archival URS data collected from mental health consumers in the state of Utah were also examined in order to answer questions regarding the validity of the measure as well as the measure's ability to detect clinically meaningful changes among consumers receiving mental health treatment.

This chapter will explore the results of this study and potential interpretations of those results. In addition, possible implications of the observed relationship between recovery as measured by the URS and distress as measured by the OQ-45 will also be explored. This chapter will proceed in the following order: reliability, validity, clinical implications, and limitations.

Reliability

In order to explore the utility of the URS as an evaluative measure it was first necessary to explore the reliability of this measure, or in other words to answer the question of whether the URS is consistent in its measurements. Reliability refers to how precise a test is, or the degree to which test scores are consistent and repeatable (American Psychological Association, American Educational Research Association, & National Council on Measurement in Education, 1985). Since the measure in question in the present study is thought to measure dynamic variables, internal consistency reliability is considered the appropriate method for estimating reliability (Ghiselli, Campbell, & Zedeck, 1981). The obtained internal consistency estimate of $\alpha = 0.898$ ($n = 91$) provides initial evidence for the reliability of the URS as it is above the .8 standard at

which measures are thought to be reliable. In other words, there is initial evidence that the URS is consistent in its measurements.

Validity

The Validity of the URS was examined in terms of its construct validity, sensitivity and specificity, sensitivity to change, and validity for change.

Construct validity. In addition to reliability, it is also important to explore an evaluative measure's validity. In other words, does the measure sufficiently measure what it purports to measure? In this case, is the URS able to measure recovery? More specifically, we were interested in exploring the construct validity of the URS, which is the degree to which an instrument accurately measures the theoretical construct that it is designed to measure.

In this study we examined evidence for the construct validity of the URS by first exploring the criterion related validity of the URS. Theoretically, a high score on a recovery measure should have a moderate negative correlation with consumer distress. Criterion-related validity is determined by how well a test score predicts some type of criterion or human behavior. In the case of recovery the criterion we are using is both consumer distress and consumer quality of life as they are measured by the OQ-45.

Initial analyses indicate a strong inverse relationship between the URS and the OQ-45 measures ($r = -0.75$, $n = 8,483$). This strong relationship provides evidence that consumers with higher distress scores are likely to have lower recovery scores and vice versa. In other words consumers that have high recovery scores are likely to report low levels of distress. This relationship provides some evidence for the construct validity of the URS in the sense that recovery and distress are thought to be inversely related. This conclusion is also supported by the URS's strong relationship to the Symptom Distress subscale of the OQ-45 ($r = -.711$, $n =$

7878). However, this stronger than expected relationship may also provide evidence that the constructs of recovery and consumer distress actually overlap to an extent that measuring recovery separately may be unnecessary. In addition, this relationship may also provide evidence that consumers receiving recovery-focused treatments are more likely to show improvement on the OQ-45. However, since these relationships are correlational, more research is needed in order to further explore these possibilities. In addition, the URS's demonstrated strong relationship to the OQ-45 subscale of Interpersonal Relations ($r=-.746$, $n=8162$) and moderate relationship to the OQ-45 subscale of Social Role ($r=-.576$, $n=8197$) provide further evidence for the construct validity for the URS by demonstrating that consumers exhibiting higher levels of recovery exhibit lower levels of relationship problems and lower levels of role dissatisfaction.

In addition, a strong inverse relationship was shown between the reverse scored quality of life items (higher scores indicate higher levels of distress) and the URS (higher scores indicate higher levels of recovery) ($r=-.863$, $n=8,216$). This strong relationship indicates that recovery seems to be highly related to consumer quality of life as expected. This strong relationship provides initial evidence for the possible link between recovery and factors that may contribute to well-being, as it is discussed in the positive psychology literature since the quality of life items ask questions which address perceived well-being, engagement and social relationships.

Sensitivity and specificity. It is also important for the construct validity of this type of measure that the measure accurately identify consumers as being in a clinical or non-clinical population (Glaros & Klein, 1988). For this reason it was important to explore the sensitivity and specificity of the URS. Sensitivity is the capacity of a measure to yield a positive result for a person with a clinical condition of interest (in this case whether a consumer of mental health

services is accurately identified as being part of a clinical population). In the case of the URS it would be the number of individuals that fall below the clinical cutoff score of 29. Similarly, specificity refers to the capacity of a measure to accurately identify someone who is not a consumer as not falling within a clinical population (Yerushalmy, 1947). In the case of the URS it would be non-consumers who score over the cutoff score of 29. In general, as the ability of the test to discriminate diagnostic groups of interest increases, so do its sensitivity and specificity.

The URS was generally accurate in both its identification of actual consumers (sensitivity = .89) as well as its identification of actual non-consumers (specificity = .95). The ability of the URS to identify both patients and non-patients suggests utilizing a cutoff score could be used to help identify when consumers move from a clinical population to a non-clinical population throughout the course of treatment. The ability of the URS to capture these differences contributes to the construct validity of the URS because the measure differentiates consumers approaching recovery from consumers who are not. The PPP of .99 indicates that of the cases the cutoff score predicted to be patients, 99% of them were actual patients, and the NPP of .04 indicates that of the cases the cutoff score predicted to be non-patients, 14% of them were actual non-patients. These indices are less useful in psychometric evaluation because they are largely contingent on the sample sizes used to calculate them (Streiner, 2003). In other words, because the clinical sample used was so much larger than the non-clinical sample, the positive and negative predictive power for the URS are less meaningful.

Given the conceptualization of recovery as a life-long process, it may be helpful to think of the cut score for the URS in terms of individuals scoring closer to one type of population (consumer) than another (non-consumers) rather than in terms of individuals possessing a specific diagnostic characteristic or disorder. In this way the cut score may be useful to

clinicians in order to help them identify consumers that are experiencing aspects of recovery that help them live meaningful, independent lives that are similar to individuals not receiving treatment.

Sensitivity to change. In addition to criterion-related validity, it is also important for an evaluative measure intended to track consumer progress in treatment to have the ability to detect consumer changes throughout the course of treatment. This is because a measure may be a valid indicator of a characteristic without being able to measure change in that characteristic (Lambert et al., 1994). Kazdin (1992) defined sensitivity for change as the ability of a dependent measure to be sensitive to the type and magnitude of change that the investigator is expecting.

In order for there to be evidence of the construct validity of the URS as an evaluative instrument it must register score changes when consumers' health statuses improve or deteriorate. For this reason it was necessary to explore whether the URS was sensitive to changes among consumers in treatment. In order to investigate the URS's ability to detect change, the URS was compared against 0 (no change). As a whole, the URS was demonstrated to be significantly different from zero when tracking consumers in treatment over time. The observed changes also occurred in the theoretically specified direction (i.e., consumers in treatment improved over time).

In terms of the individual URS items, all 10 of the items demonstrated change in the theoretically specified direction in that they had positive slopes; however, only seven of the 10 items demonstrated a rate of change that was statistically significant when compared to zero. This finding provides evidence that both the URS as a whole the seven items can detect changes in consumers in treatment over time. The three items that were not demonstrated to be able to

detect change when compared to zero were “I have a place to live and it’s ok,” “I have people/friends I can turn to,” and “I have meaningful relationships.”

There are a number of possibilities that may have contributed to the inability of these items to detect changes in consumers that were significantly different from zero over time. One possibility is that these questions are poor indicators of aspects of consumers’ lives that are likely to change during mental health treatment. Another possibility is that outside consumer relationships and living situations are not emphasized enough in the mental health treatment programs in Utah for there to be measurable changes made by consumers in those areas. A third possibility is that the areas that these items measure may change more slowly over time than the other areas measured by the URS. In other words, if we tracked consumers for a longer period of time perhaps significant changes in these areas might be observed. Yet another possibility could be that these items cover areas that may change once but not gradually over time (i.e., once someone has a place to live that they are ok with they are not likely to have this progress any more).

One potential way to improve the sensitivity to change of the URS would be to remove these items; however, if the inability of these items to detect changes that are significantly different from no change is due to a lack of emphasis in these recovery areas in consumer treatment removing these items would only serve to further take attention away from these recovery areas deemed important by consumers. For this reason, further inquiry is needed in order to ascertain the likely cause of the failure of these items to detect changes in consumers before a final decision can be made about whether to remove the items from the measure.

Validity for change. In addition to requiring an evaluative measure such as the URS to be demonstrated as being sensitive to change, it is also important to examine the validity of the

changes that are measured by the instrument. For this reason the consumer treatment trajectories as measured by the URS were compared with consumer treatment trajectories as measured by the OQ-45. When HLM was used to compare consumer treatment trajectories over time between the two measures, they were found to track change in a similar way overall. When they were examined among different classifications of consumers, the measures were again found to be mostly similar. In cases where the slopes of the two measures were found to be significantly different there was no difference in the directionality of the slopes (positive vs. negative) and the URS was consistently found to have a shallower slope in such situations, indicating that the URS may not be as sensitive to certain types of changes as the OQ-45.

The observed similarities between consumer changes as measured by the URS and consumer changes as measured by the OQ-45 provide evidence for the URS's validity for change in the sense that recovery and distress are theoretically related to one another. Further, the observed differences may suggest that the two domains are somewhat different in terms of the way consumers change throughout the course of treatment. One possible explanation for the observed differences between the two measures could be that consumer changes in the domain of recovery occur more gradually than consumer changes in the domain of distress. For this reason it may be more useful to administer recovery instruments, such as the URS, less frequently than distress instruments, such as the OQ-45. Such an approach may also be practically beneficial so as not to overly burden consumers with the frequent administration of too many outcome measures throughout the course of their treatment.

Another area to explore in terms of examining an evaluative measure's validity for measuring change is whether the measure is able to detect clinically significant changes in consumers over time (Jacobson, Truax, and Kazdin, 1992). The ability of the URS to detect

clinically significant changes was examined by calculating an RCI for the measure and then comparing this with the established RCI of the OQ-45 (Jacobson et al., 1992). Calculating an RCI for the URS allowed us to examine whether consumers were making observable changes that show clinically useful movement from a clinical population towards a non-clinical population. In other words, it allowed us to explore whether the URS might be able to detect changes in consumers that might be useful to clinicians that are attempting to determine the efficacy of treatment. Utilizing the RCI, the URS identified 12% of consumers as recovered (scoring within a range that is most likely to be in a non-clinical population), 77% of consumers as showing no change, 5% of consumers as improved (moving significantly towards scores that are likely to fall in a non-clinical population), and 6% of consumers as deteriorated (moving significantly away from scores that are likely to fall within a non-clinical population).

The RCI classifications observed among consumers by the URS were then compared with RCI classifications made by the OQ-45. In terms of change metrics the two measures appear to be capturing somewhat different phenomena. When compared, the URS identified the proportion of consumers in the same category as the OQ-45 58% of the time. However, when broken down into individual groups the URS change categorizations were equal to the OQ-45 to a much lesser extent with the URS identifying only 42% of the consumers classified as recovered by the OQ-45, 12% of the consumers classified as improved by the OQ-45, and 23% of the consumers classified as deteriorated by the OQ-45. In all cases, with the exception of the no change classification, the URS identified fewer consumers in each category than the OQ-45. This fits with the finding that recovery as measured by the URS may be less sensitive to clinical changes than distress as measured by the OQ-45. These results provide further evidence that

distress measured during the course of clinical treatment changes more drastically than recovery. For this reason, the URS may be most useful as an occasional measure of treatment outcome.

Clinical Implications

It is important to discuss several possible clinical implications of the results of this study. The first is that initial evidence for the reliability, validity, sensitivity to change, and validity for change suggest that the URS may be a useful instrument for clinicians and community mental health programs which are seeking to measure recovery among consumers over time. Given the results of the study, the URS appears to be most useful as an occasional evaluative measure of recovery. In other words, since changes in recovery as measured by the URS appear to be less drastic, it is likely that the URS would most likely be useful as an occasional evaluative measure (i.e., to be administered every month rather than every clinical visit or prior to case conferences etc.).

The second implication is that the strong inverse relationship observed between recovery and distress is encouraging for community mental health centers that may be implementing recovery-oriented interventions. This relationship suggests that the implementations of recovery-oriented interventions are unlikely to detract from treatment efforts that are aimed at reducing consumer symptoms or distress. Further the strength of the relationship between the two measures may even suggest that recovery-oriented interventions may bolster consumer gains in distress reduction. It is important to note that, since this study was not experimental, further inquiry is needed in order to fully explore the nature of the observed relationship between recovery and consumer distress.

The third implication is that the strong relationship between the quality of life items from the OQ-45 and the URS also provides evidence for the potential relationship between recovery

and well-being as it has been discussed in the positive psychology literature. For this reason, it might be helpful for clinicians to look towards positive psychology research, particularly in areas of engagement, reported well-being, and social relationships, as a way of developing and implementing recovery oriented interventions.

Limitations

In order for an evaluative measure to be valid for measuring changes in consumers over time it must also measure stability in consumers who are not receiving treatment when compared to consumers who are. For this reason, one of the major remaining questions regarding the validity of the URS is test-retest reliability. A high test-retest reliability estimate observed in a group of consumers not receiving treatment would provide evidence that the URS is actually detecting changes that are the result of mental health treatment. Further, multiple administrations given to a group of consumers not receiving treatment could also be compared against consumers receiving treatment to further explore whether the URS as a measure and its individual items are sensitive to change over time. This study compared the URS and the individual items on the URS to zero; however, comparing it to a sample of individuals not in treatment would provide better evidence for the sensitivity of the URS for change because we know that even without treatment consumers may show a small amount of improvement over time (Aneshensel, Estrada, Hansell, & Clark, 1987; Bromet, Dunn, Connell, Dew, & Schulberg, 1986; Durham, Burlingame, & Lambert, 1998; Henderson, Byrne, & Duncan-Jones, 1981; Jorm, Duncan-Jones, & Scott, 1989).

An additional limitation was that consumers were not given the URS at intake. Because of this we were unable to examine consumer treatment trajectories that started at intake. We

were able to look at consumers beginning treatment in comparison to the OQ-45, but most often these consumers were not given the URS until their second or third session. Examining a group of consumers taking the URS at intake and then tracking them throughout treatment might give us not only a more clear picture of recovery's longitudinal relationship to distress, but also might allow us to examine any time differences between the two (e.g., perhaps a certain level of distress reduction is required before change in recovery can be observed).

It is also important to note that, since this study was not experimental, no direct conclusions can be made about causality in terms of recovery or the relationship between recovery-oriented interventions and the reduction of consumer distress due to treatment. For this reason it is important to note that the observed relationship between these two measures could be related to other unspecified variables. The results of this study do, however, provide a rationale for conducting an experimental study comparing consumers receiving recovery-oriented interventions with consumers that are receiving other treatments.

Lastly, it is also important to note that while the URS was developed utilizing consumer feedback and focus groups, it was not constructed utilizing a large pool of items. This is important to note because it is possible that utilizing a large pool of items in its construction would have provided a more conservative approach to insuring that the domain of recovery was adequately captured by the measure. For this reason, further exploration is needed in comparing the URS with other measures of recovery that are thought to cover multiple domains of recovery such as the RAS (Corrigan et al., 1999).

Future Directions

In order to further bolster the construct validity of the URS it may be helpful to explore whether there is evidence for the temporal stability of the URS. In other words, a test-retest

reliability estimate would further add to the evidence as to whether the URS is a useful tool for measuring consumer changes according to the construct of recovery. Obtaining multiple administrations from a group of consumers not receiving treatment would also provide a means for exploring whether the changes measured by the URS are actually the result of consumers receiving mental health treatment.

In addition, in order to further provide evidence for the construct validity of the URS, it would be helpful to compare the URS with other established recovery instruments, such as the RAS or the STORI (Andresen et al., 2006; Corrigan et al., 1999). While these instruments have not been explored in terms of their ability to detect change, they provide additional ways of measuring the construct of recovery that could prove useful in examining whether the URS adequately captures the construct of recovery.

As previously mentioned, a direct comparison study is needed to explore differences in outcomes between consumers receiving treatments based on the recovery model and consumers that are receiving other types of treatments. This would allow for a more specific exploration of the strong inverse relationship between recovery and distress that was observed in this study. Further exploring the specifics of this relationship would allow clinicians to gain a better understanding of the types of interventions that contribute to both an increase in consumer recovery and a decrease in consumer distress.

The strong relationship between the URS and the quality of life items on the OQ-45 also provides initial evidence that there may be a relationship between recovery and well-being. In order to further explore this observed relationship it would be necessary to compare a recovery measure such as the URS with a gold standard measure of well-being. This comparison would allow for further exploration of the relationship between recovery and well-being. In addition, a

controlled study that utilizes positive psychology interventions would be able to explore more specifically whether well-being oriented interventions lead to an increase in consumer recovery.

References

- Aiken, L. R. (1977). Note on sensitivity: A neglected psychometric concept. *Perceptual and Motor Skills*, 45(3), 1330-1330.
- Allen, M., & Yen, W. (1979). *Introduction to measurement theory*. Monterey, California: Brooks/Cole Publishing Company.
- American Psychological Association., American Educational Research Association., & National Council on Measurement in Education. (1985). *Standards for Educational & Psychological Testing*. Hyattsville, MD: American Psychological Association.
- Aneshensel, C. S., Estrada, A. L., Hansell, M. J., & Clark, V. A. (1987). Social psychological aspects of reporting behavior: Lifetime depressive episode reports. *Journal of Health and Social Behavior*, 28, 232–246.
- Andresen, R., Caputi, P., & Oades, L. (2006). Stages of Recovery Instrument: Development of a measure of recovery from serious mental illness. *Australian and New Zealand Journal of Psychiatry*, 40(11), 972-980.
- Anthony, W. A. (2003). Expanding the evidence base in an era of recovery. *Psychiatric Rehabilitation Journal*, 27(1), 1-2.
- Arnold, C. (1992). An introduction to hierarchical linear models. *Measurement & Evaluation in Counseling & Development*, 25(2), 58.
- Bromet, E. J., Dunn, L. O., Connell, M. M., Dew, M. A., & Schulberg, H. C. (1986).

- Long-term reliability of diagnosing lifetime major depression in a community sample. *Archives of General Psychiatry*, 43, 435–440.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 101(1), 147-158.
- Burlingame, G. M., Lambert, M. J., Reisinger, C. W., & Neff, W. M. (1995). Pragmatics of tracking mental health outcomes in a managed care setting. *Journal of Mental Health Administration*, 22(3), 226-236.
- Campbell-Orde, T., Chamberlin, J., Carpernter, J., & Leff, H. S. (2005). *Measuring the promise: A compendium of recovery measures* (Vol. 2). Cambridge, MA: The Evaluation Center at the Human Services Research Institute.
- Compton, W. C. (2005). *Introduction to Positive Psychology*. Belmont, CA, US: Thomson Wadsworth.
- Corrigan, P. W., Giffort, D., Rashid, F., Leary, M., & Okeke, I. (1999). Recovery as a psychological construct. *Community Mental Health Journal*, 35(3), 231-239.
- Corrigan, P. W., Salzer, M., Ralph, R. O., Sangster, Y., & Keck, L. (2004). Examining the factor structure of the Recovery Assessment Scale. *Schizophrenia Bulletin*, 30(4), 1035-1041.
- Davidson, L., Lawless, M. S., & Leary, F. (2005). Concepts of recovery: Competing or complementary? *Current opinion in psychiatry*, 18(6), 664-667.
- Davidson, L., O'Connell, M., Tondora, J., Styron, T., & Kangas, K. (2006). The top ten concerns about recovery encountered in mental health system transformation. *Psychiatric Services (Washington, D.C.)*, 57(5), 640-645.

Davidson, L., O'Connell, M. J., Tondora, J., Lawless, M., & Evans, A. C. (2005). Recovery in serious mental illness: A new wine or just a new bottle? *Professional*

Psychology: Research and Practice, 36(5), 480-487.

- Davidson, L., & Roe, D. (2007). Recovery from versus recovery in serious mental illness: One strategy for lessening confusion plaguing recovery. *Journal of Mental Health, 16*(4), 459-470.
- Deyo, R. A., Diehr, P., & Patrick, D. L. (1991). Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Controlled Clinical Trials, 12*(4 Suppl), 142S-158s.
- Durham, C., Burlingame, G. M., & Lambert, M. J. (1998). *Outcome Questionnaire: Repeated administrations, mechanical responding, and social desirability*. Unpublished manuscript, Department of Psychology, Brigham Young University, Provo, UT.
- Ellsworth, J. R., Lambert, M. J., & Johnson, J. (2006). A comparison of the outcome Questionnaire-45 and Outcome Questionnaire-30 in classification and prediction of treatment outcome. *Clinical Psychology & Psychotherapy, 13*(6), 380-391.
- Fisher, D. B., & Ahern, L. (2002). Evidence-based practices and recovery. *Psychiatric Services, 53*(5), 632-633.
- Fowler, R. D., Seligman, M. E. P., & Koocher, G. P. (1999). The APA 1998 annual report. *American psychologist, 54*(8), 537-568.
- Fredrickson, B. L., & Losada, M. F. (2005). Positive affect and the complex dynamics of human flourishing. *American Psychologist, 60*(7), 678-686.
- Frese, F. J., III, Stanley, J., Kress, K., & Vogel-Scibilia, S. (2001). Integrating evidence-based practices and the recovery model. *Psychiatric Services, 52*(11), 1462-1468.
- Froh, J. J. (2004). The history of positive psychology: Truth be told. *NYS Psychologist, 16*(3), 18-20.
- Gable, S. L., & Haidt, J. (2005). What (and why) is positive psychology? *Review of*

- General Psychology*, 9(2), 103-110.
- Ghiselli, E., Campbell, J., & Zedeck, S. (1981). *Measurement Theory for the Behavioral Sciences*. San Francisco, CA: W.H. Freeman and Company.
- Glaros, Alan G.; Kline, & Rex B. (1988). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. *Journal of clinical psychology*, 44(6), 1013-1023.
- Griner, Mayewski, Mushlin, & Greenlan. (1981). Selection and interpretation of diagnostic tests and procedures. *Annals of Internal Medicine*, 94, 553-592.
- Guyatt, G., Walter, S., & Norman, G. (1987). Measuring change over time: assessing the usefulness of evaluative instruments. *Journal Of Chronic Diseases*, 40(2), 171-178.
- Hatfield, D. R., & Ogles, B. M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice*, 35(5), 485-491.
- Henderson, A. S., Byrne, D. G., & Duncan-Jones, P. (1981). *Neurosis and the Social Environment*. New York: Academic.
- Jacobson, N. (2004). In *recovery: The Making of Mental Health Policy*. Nashville, Tennessee: Vanderbilt University University Press.
- Jacobson, N., & Curtis, L. (2000). Recovery as policy in mental health services: Strategies emerging from the states. *Psychiatric Rehabilitation Journal*, 23(4), 333-341.
- Jacobson, N., & Greenley, D. (2001). What is recovery? A conceptual model and explication. *Psychiatric Services*, 52(4), 482-485.
- Jacobson, N.S., Follette, W.C., & Revenstorff, D. (1984). Psychotherapy outcome

- research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15, 336-452.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12-19.
- Jacobson, N. S., Truax, P., & Kazdin, A. E. (1992). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. In *Methodological issues & strategies in clinical research*. (pp. 631-648). Washington, DC, US: American Psychological Association.
- Jacobson, N.S., Follette, W.C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15, 336-452.
- Jerrell, J. M., Cousins, V. C., & Roberts, K. M. (2006). Psychometrics of the Recovery Process Inventory. *The Journal of Behavioral Health Services & Research*, 33(4), 464-473.
- Jorm, A. F., Duncan-Jones, P., & Scott, R. (1989). An analysis of the re-test artifact in longitudinal studies of psychiatric symptoms and personality. *Psychological Medicine*, 19, 487-493.
- Kazdin, A. E. (1998). *Research design in clinical psychology* (3rd ed.). Needham Heights, MA US: Allyn & Bacon.
- Keyes, C. L. M. (2007). Promoting and protecting mental health as flourishing: A complementary strategy for improving national mental health. *American Psychologist*, 62(2), 95-108.

- Kirshner, B., & Guyatt, G. (1985). A Methodological framework for assessing health indices. *Journal of chronic diseases*, 38(1), 27-36.
- Lambert, M. J., Gregersen, A. T., Burlingame, G. M., & Maruish, M. E. (2004). The Outcome Questionnaire-45. In *The use of psychological testing for treatment planning and outcomes assessment: Volume 3: Instruments for adults (3rd ed)*. (pp. 191-234). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Lambert, M. J., Hill, C. E., Bergin, A. E., & Garfield, S. L. (1994). Assessing psychotherapy outcomes and processes. In *Handbook of psychotherapy and behavior change (4th ed.)*. (pp. 72-113). Oxford England: John Wiley & Sons.
- Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy research*, 11(1), 49-68.
- Lipsey, M. W. (1983). A scheme for assessing measurement sensitivity in program evaluation and other applied research. *Psychological bulletin*, 94(1), 152-165.
- O'Connor, F. W., & Delaney, K. R. (2007). The recovery movement: Defining evidence-based processes. *Archives of Psychiatric Nursing*, 21(3), 172-175.
- Ogles, B. M., Lambert, M. J., & Fields, S. A. (2002). *Essentials of outcome assessment*. Hoboken, NJ, US: John Wiley & Sons Inc.
- Okiishi, J., Lambert, M. J., Eggett, D., Nielsen, L., Dayton, D. D., & Vermeersch, D. A. (2006). An Analysis of Therapist Treatment Effects: Toward Providing Feedback to Individual Therapists on Their Clients' Psychotherapy Outcome. *Journal of Clinical Psychology*, 62(9), 1157-1172.

- Okiishi, J., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An empirical analysis of therapist effects. *Clinical Psychology & Psychotherapy, 10*(6), 361-373.
- Peysers, H. (2001). Commentary: What is recovery? A commentary. *Psychiatric Services, 52*(4), 486-487.
- Peysers, H. S. (2001). 'A conceptual model of recovery': Reply. *Psychiatric Services, 52*(5), 688-689.
- Raudenbush, S. W., & Chan, W.S. (1993). Application of a hierarchical linear model to the study of adolescent deviance in an overlapping cohort design. *Journal of consulting and clinical psychology, 61*(6), 941-951.
- Resnick, S., Warmoth, A., & Selin, I. A. (2001). The humanistic psychology and positive psychology connection: Implications for psychotherapy. *Journal of Humanistic Psychology, 41*(1), 73-101.
- Resnick, S. G., & Rosenheck, R. A. (2006). Recovery and positive psychology: Parallel themes and potential synergies. *Psychiatric services, 57*(1), 120-122.
- Roe, D., & Kravetz, S. (2003). Different ways of being aware of a psychiatric disability: A multifunctional narrative approach to insight into mental disorder. *Journal of nervous and mental disease, 191*(7), 417-424.
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of personality and social psychology, 57*(6), 1069-1081.
- Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of personality and social psychology, 69*(4), 719-727.

- Sawyer, K. (2002). Toward a Positive Psychology. *Psychcritiques*, 47(6), 691-693.
- Seligman, M. E. P. (1998). The president's address. *American Psychologist*, 1998 Annual Report, 559-562.
- Seligman, M. E. P., Peterson, C., Aspinwall, L. G., & Staudinger, U. M. (2003). Positive clinical psychology. In A psychology of human strengths: Fundamental questions and future *directions for a positive psychology*. (pp. 305-317). Washington, DC, US: American Psychological Association.
- Seligman, M. E. P., Steen, T. A., Nansook, P., & Peterson, C. (2005). Positive psychology progress: Empirical validation of interventions. *American psychologist*, 60(5), 410-421.
- Spaulding, W., & Nolting, J. (2006). Psychotherapy for schizophrenia in the year 2030: Prognosis and Prognostication. *Schizophrenia Bulletin*, 32, S94-s105.
- Speer, D. C., & Greenbaum, P. E. (1995). Five methods for computing significant individual client change and improvement rates: Support for an individual growth curve approach. *Journal of consulting and clinical psychology*, 63(6), 1044-1048.
- Substance Abuse and Mental Health Services Administration (SAMHSA), (2005). *Transforming mental health care in America. The Federal Action agenda: First Steps*. Retrieved September 5, 2007, from http://www.samhsa.gov/Federalactionagenda/NFC_TOC.aspx.
- Substance Abuse and Mental Health Services Administration (SAMHSA), (2006). *National consensus statement on mental health recovery*. Retrieved September 10, 2007, 2007, from <http://mentalhealth.samhsa.gov/publications/allpubs/sma05-4129/>
- Tingey, Lambert, Burlingame, & Hansen (1996). Assessing clinical significance:

- Proposed extensions to method. *Psychotherapy research*, 6(2), 109-123.
- Vecchio, T.J. (1966). Predictive value of a single diagnostic test in unselected populations. *N. Engl. J. Med.*, 274, 1171.
- Vermeersch, D. A. (1998). Sensitivity to Change as an Item Selection Criteria for an Outcome Measure. Unpublished Dissertation, Brigham Young University, Provo, UT.
- Vermeersch, D. A., Lambert, M. J., & Burlingame, G. M. (2000). Outcome questionnaire: Item sensitivity to change. *Journal of personality assessment*, 74(2), 242-261.
- 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(1), 38-49.
- Ware, J. H. (1985). Linear models for the analysis of longitudinal statistics. *American Statistician*, 39, 95-101.
- Wells, M. G., Burlingame, G. M., Lambert, M. J., Hoag, M. J., & Hope, C. A. (1996). Conceptualization and measurement of patient change during psychotherapy: Development of the Outcome Questionnaire and Youth Outcome Questionnaire. *Psychotherapy: Theory, Research, Practice, Training*, 33(2), 275-283.
- Yerushalmy, J. (1947). Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. *Public Health Reports*. 62, 1432-1449

Appendix A

Utah Recovery Indicators

Name: _____ Male: <input type="checkbox"/> Female: <input type="checkbox"/> ID#: _____ Date of Birth: ___/___/___ /___	Today's Date: ___/___/___
--	---------------------------

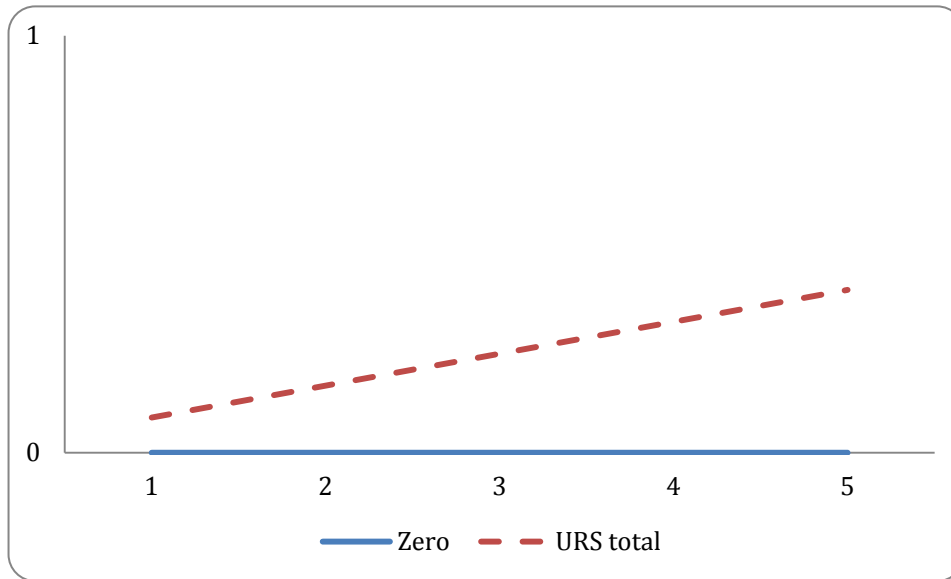
Instructions: The following questions ask about how you feel about working together in a group. Please read each question carefully, and then mark the box that best describes HOW MUCH OF THE TIME YOU HAVE FELT THIS WAY DURING THE PAST MONTH (30 days). There are no right or wrong answers, so please be as honest as you can.

	<u>Never</u>	<u>Rarely</u>	<u>Sometimes</u>	<u>Frequently</u>	<u>Almost Always</u>	Do not mark below
1. I have hope for the future.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	
2. I have meaningful work, volunteer work or activities in my life.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	
3. Most days I get to do something I enjoy.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	
4. I have a place to live and it's ok.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	
5. My life has meaning.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	
6. I have people/friends that I can turn to.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	
7. I am connected to my community.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	
8. I am in charge of my own life and recovery.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	
9. I have goals for my future.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	
10. I have meaningful relationships.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	
Total Score						

Developed by NAMI-Utah in collaboration with Brigham Young University

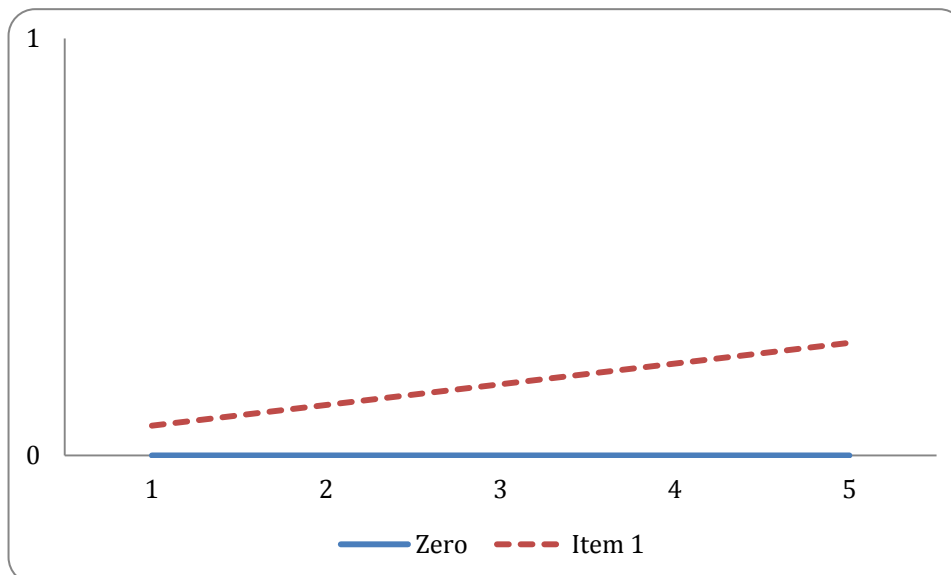
Appendix C: Individual item HLM analysis graphs

URS total score compared to zero



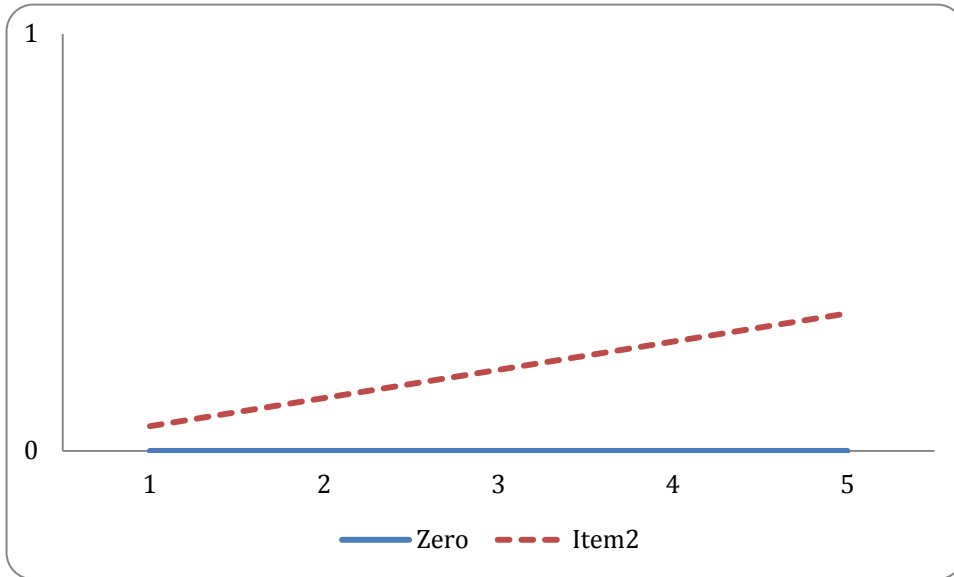
	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.07654	0.01304	1093	5.87	0.0001

URS item 1 compared to zero



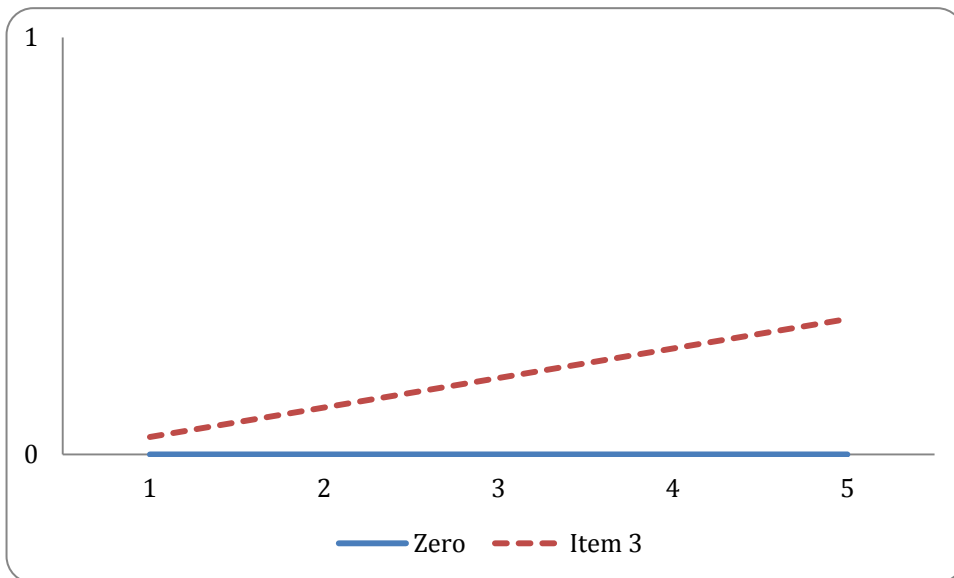
	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.04967	0.01561	1016	3.18	0.0015

URS item 2 compared to zero



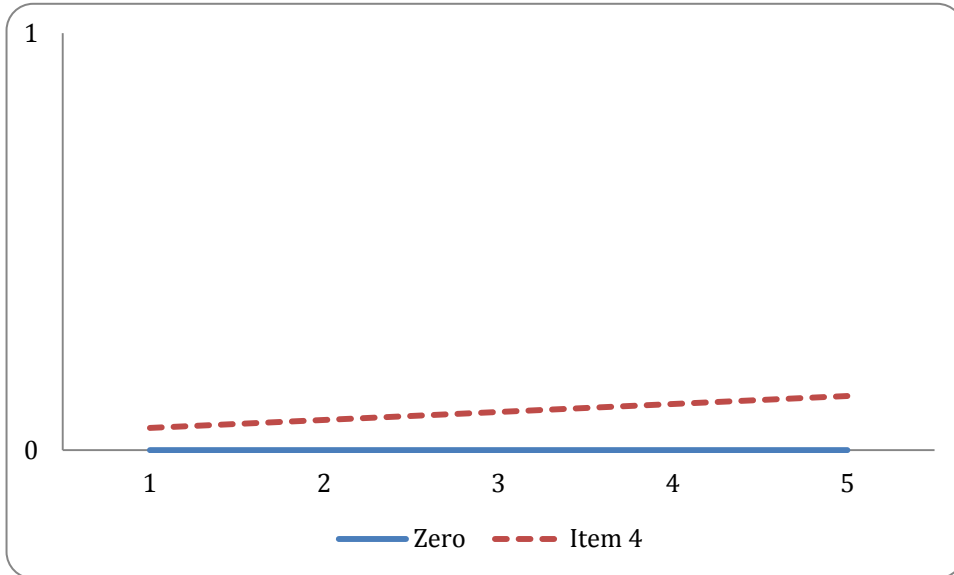
	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.06763	0.01678	1075	4.03	0.0001

URS item 3 compared to zero



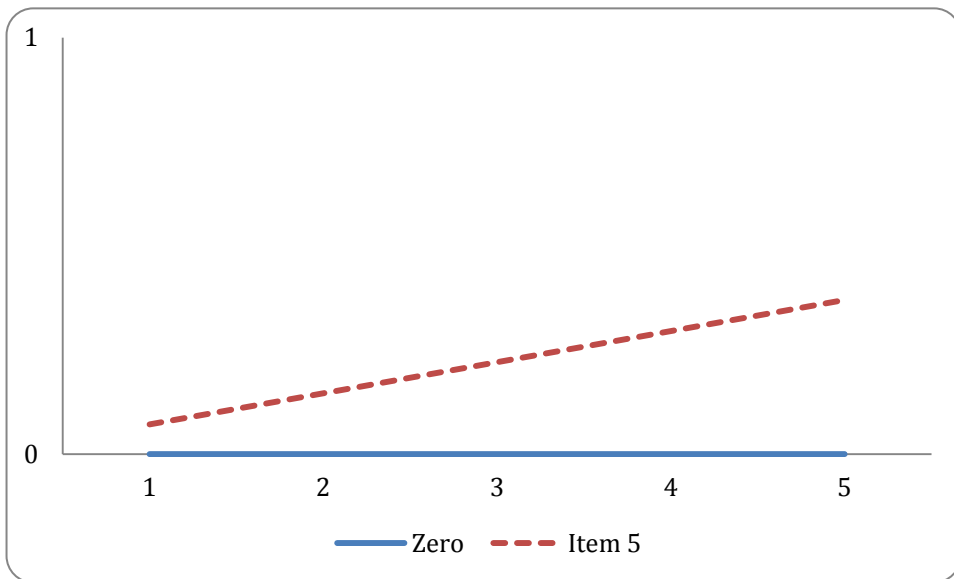
	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.07075	0.01584	1008	4.47	0.0001

URS item 4 compared to zero



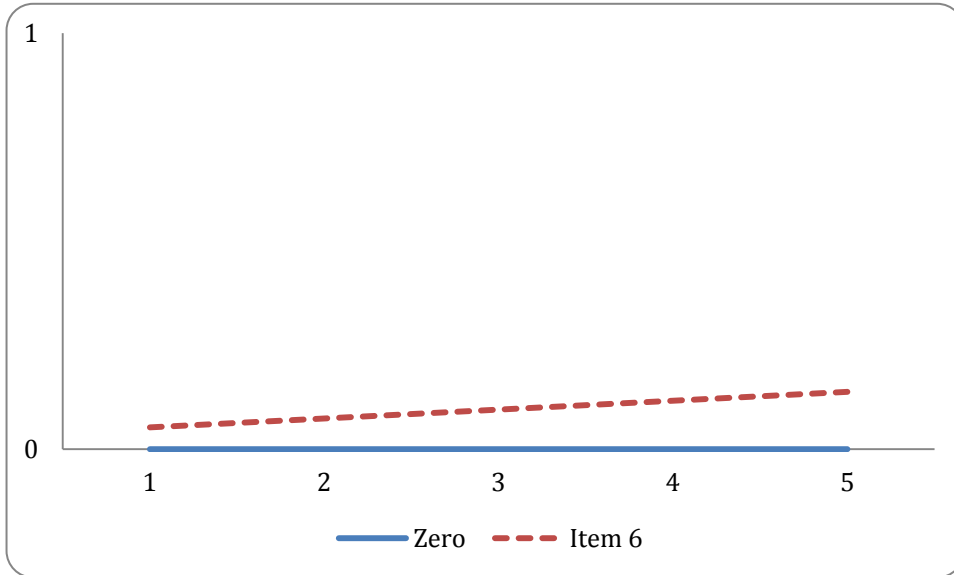
	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.01913	0.01643	1177	1.16	0.2446

URS item 5 compared to zero



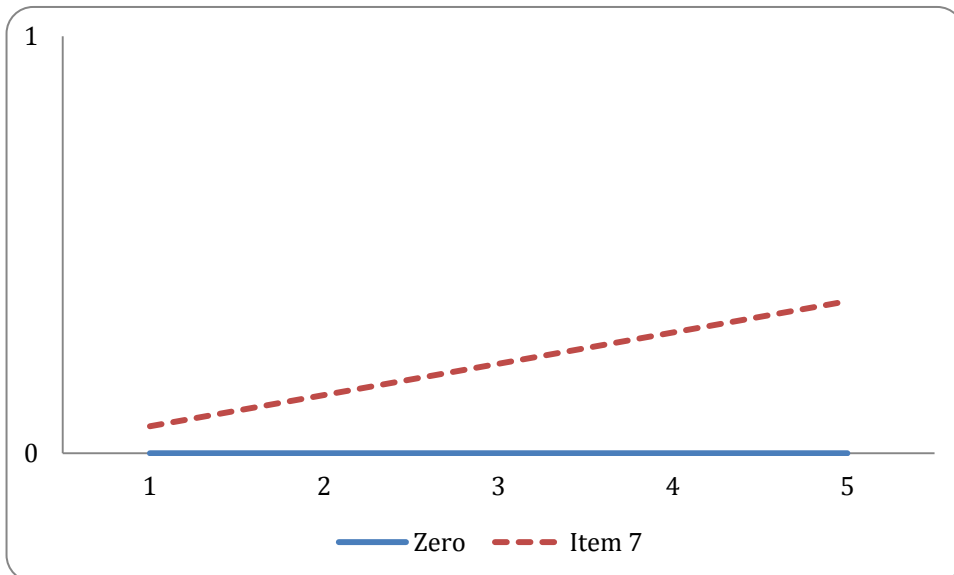
	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.0747	0.01447	916	5.16	0.0001

URS item 6 compared to zero



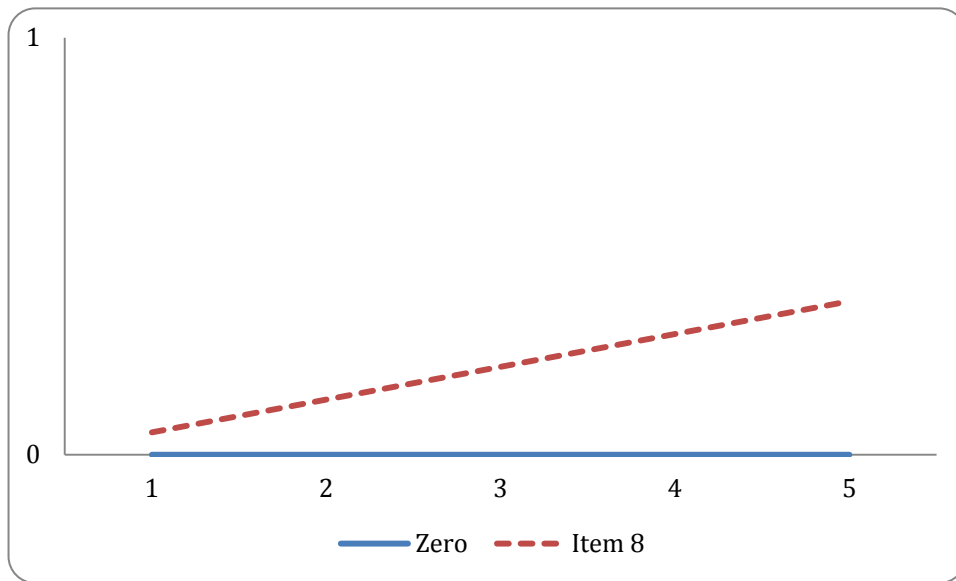
	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.02135	0.01569	1044	1.36	0.174

URS item 7 compared to zero



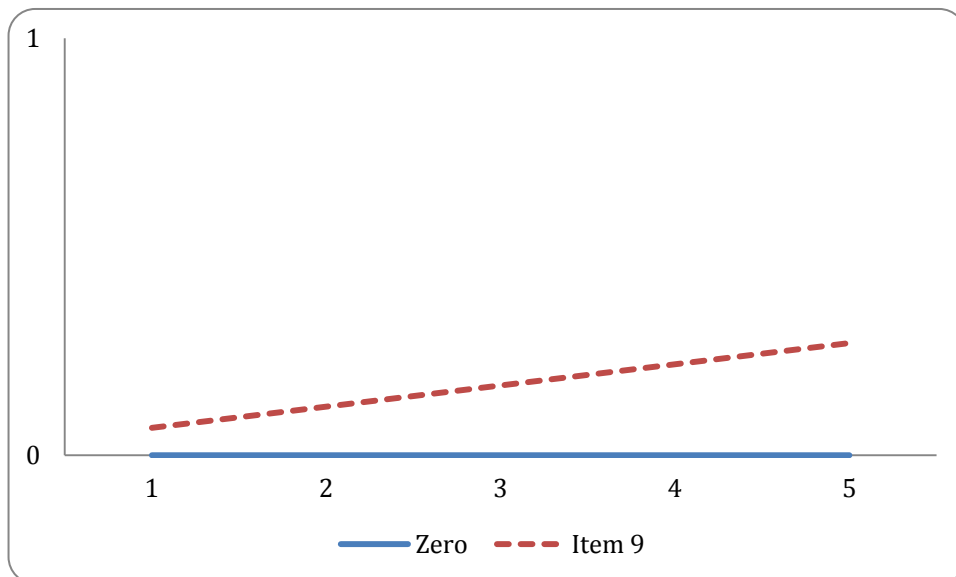
	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.075	0.01502	1053	4.99	0.0001

URS item 8 compared to zero



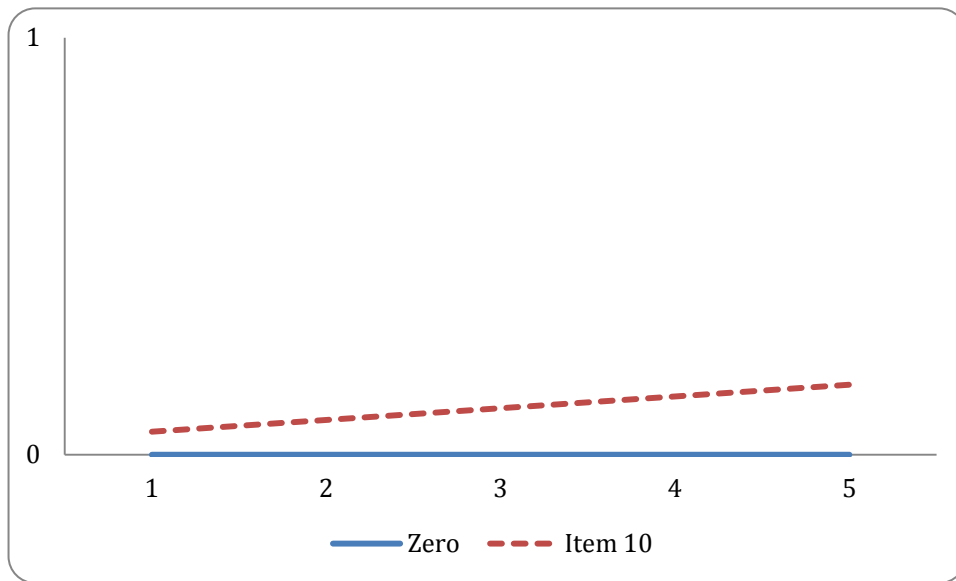
	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.07862	0.01721	1121	4.57	0.0001

URS item 9 compared to zero



	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.05082	0.01459	941	3.48	0.0005

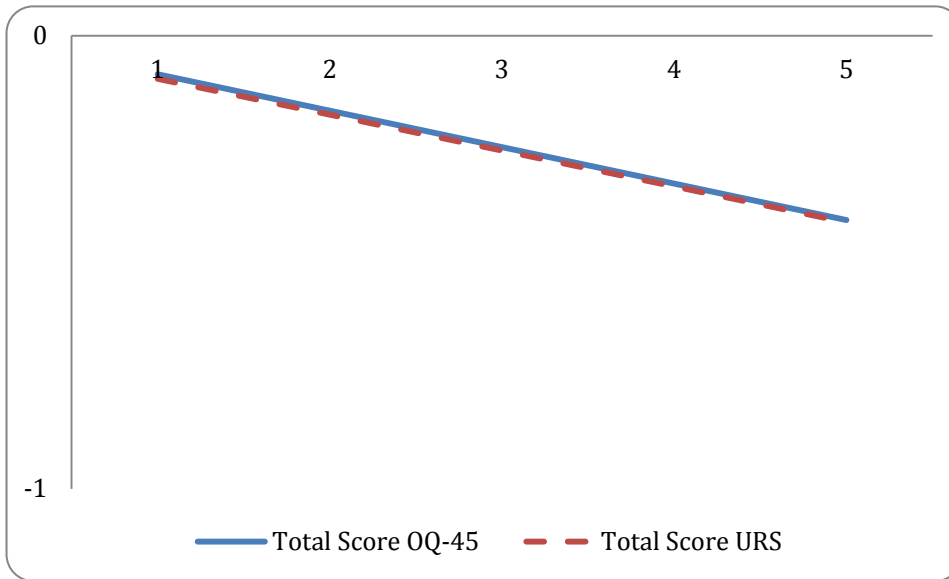
URS item 10 compared to zero



	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.02817	0.1597	1161	1.76	0.078

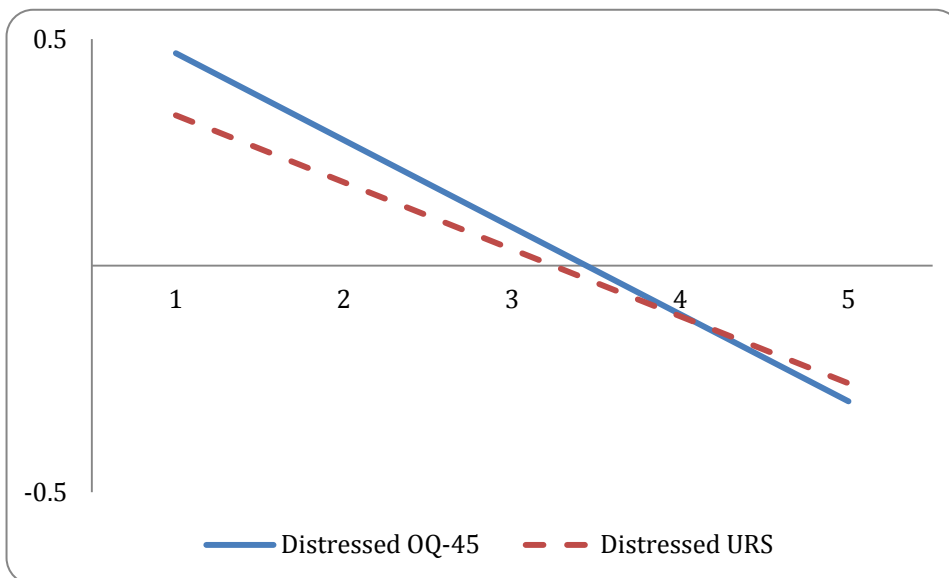
Appendix D: URS and OQ-45 HLM Treatment Trajectory Comparisons Over Time.

OQ-45 and URS total scores



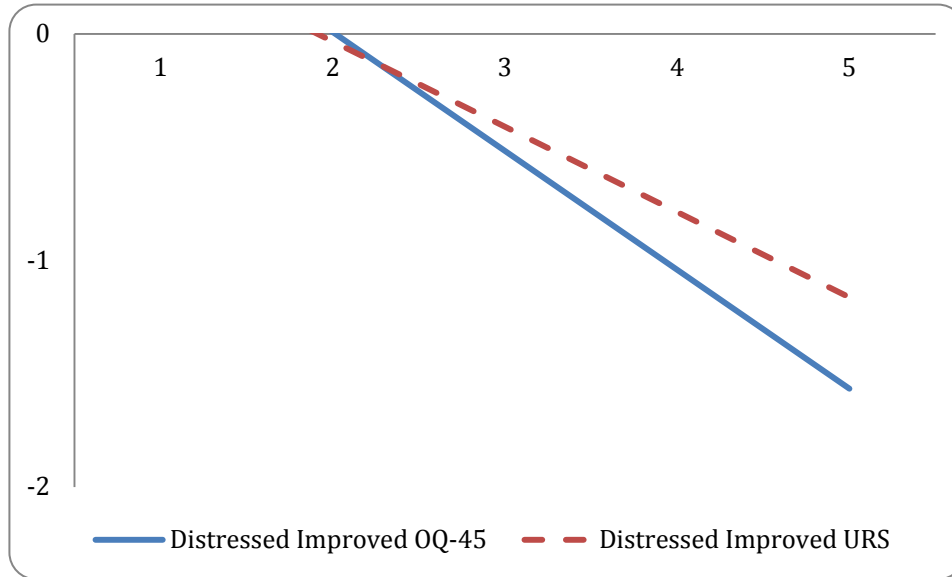
	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.001246	0.01384	9080	0.09	0.9282

Consumers categorized by the OQ-45 as distressed



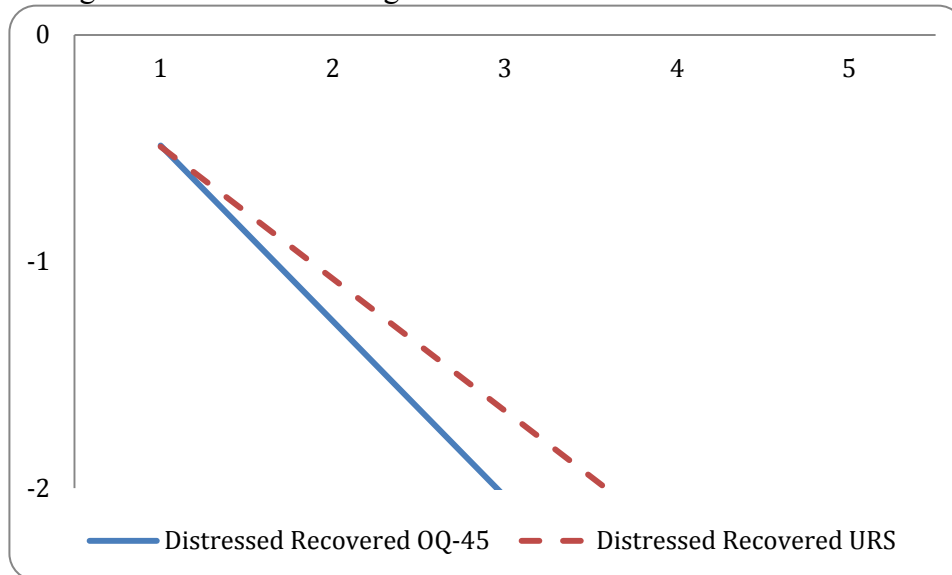
	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.04419	0.01615	5774	2.74	0.0062

Consumers categorized by the OQ-45 as starting treatment clinically distressed and improved over time.



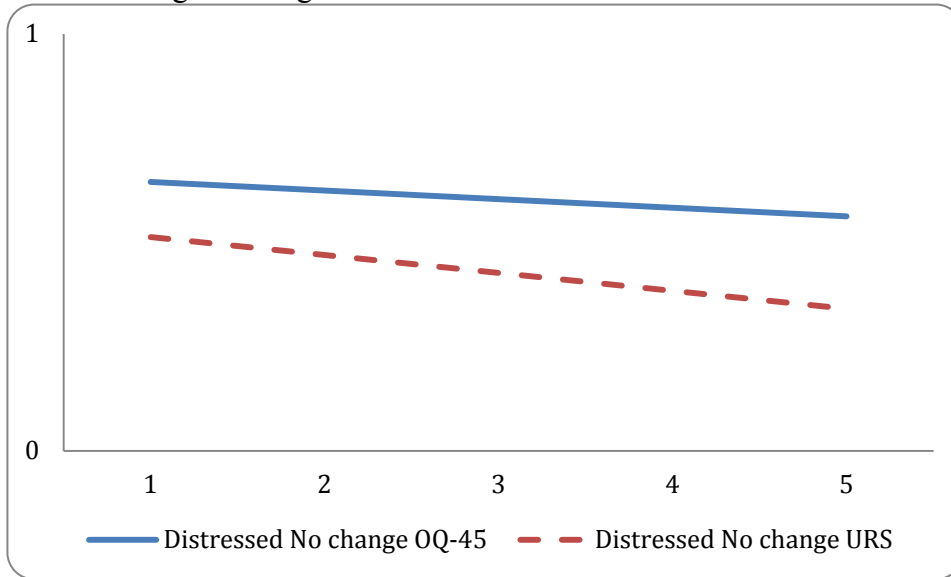
	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.1482	0.03642	1156	4.07	0.0001

Consumers categorized by the OQ-45 as starting treatment in the clinically distressed range and moving into the recovered range.



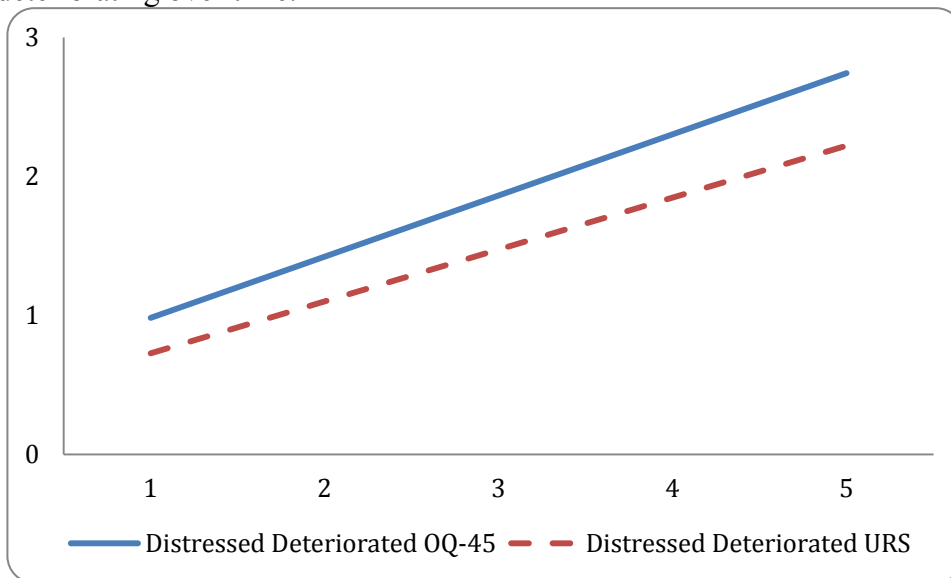
	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.192	0.04578	974	4.19	<.0001

Consumers categorized by the OQ-45 as starting treatment in the clinically distressed range and then exhibiting no change over time.



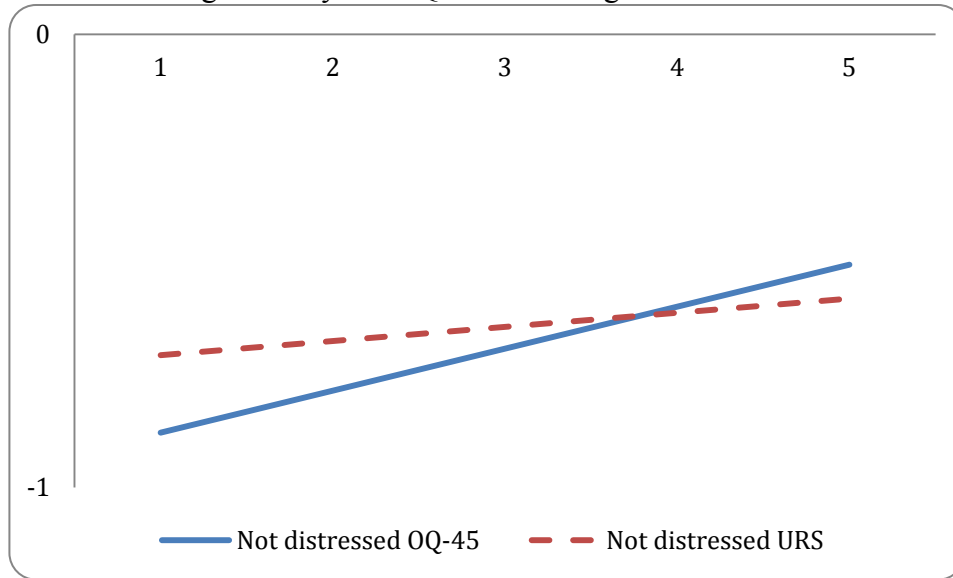
	Estimate	Standard Error	DF	T value	pr>t
Interaction	-0.02235	0.01876	3193	-1.19	0.2337

Consumers categorized by the OQ-45 as starting treatment in the distressed range and then deteriorating over time.



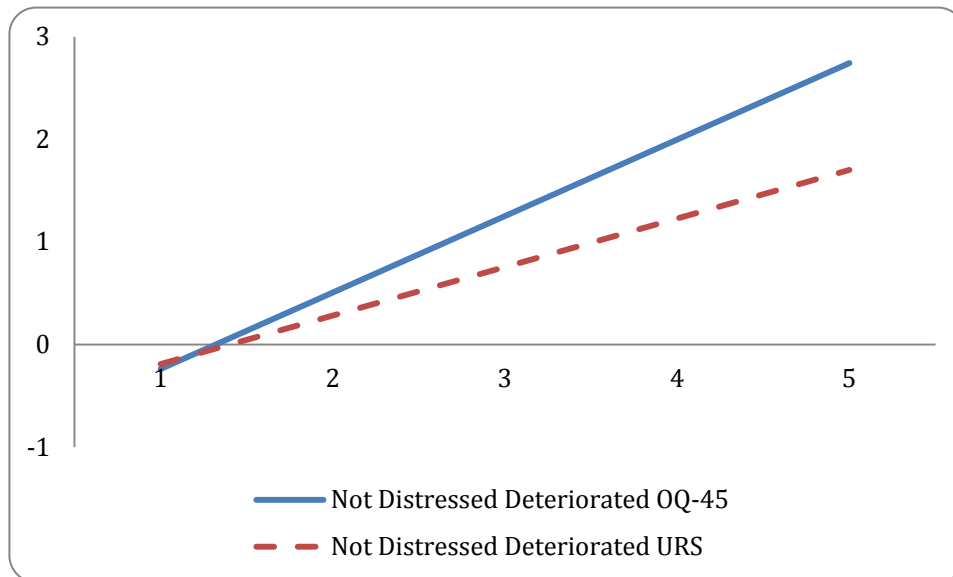
	Estimate	Standard Error	DF	T value	pr>t
Interaction	-0.06672	0.05423	107	-1.23	0.2213

Consumers categorized by the OQ-45 as starting treatment in the not distressed range.



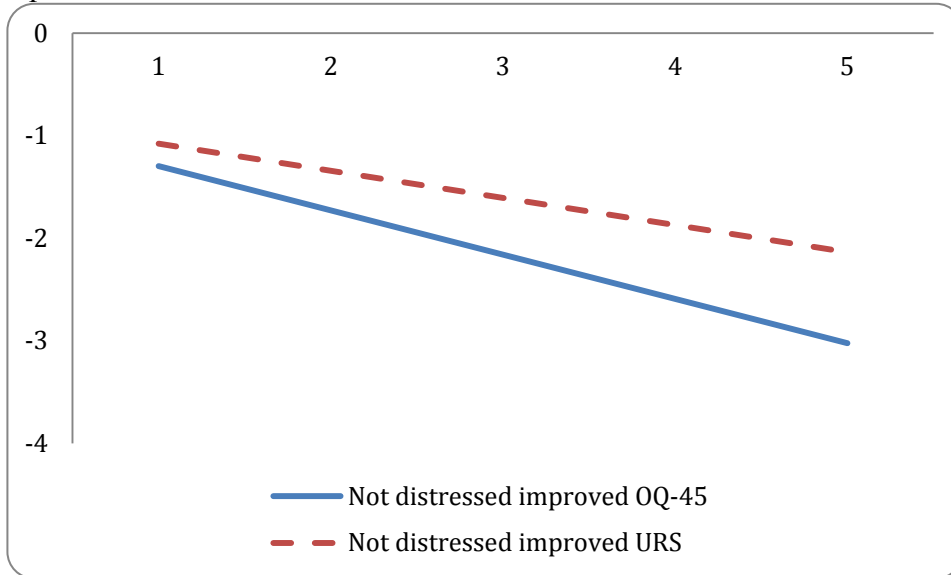
	Estimate	Standard Error	DF	T value	pr>t
Interaction	-0.06148	0.02481	415	-2.48	0.0136

Consumers categorized by the OQ-45 as starting treatment in the not distressed range and then deteriorated over time.



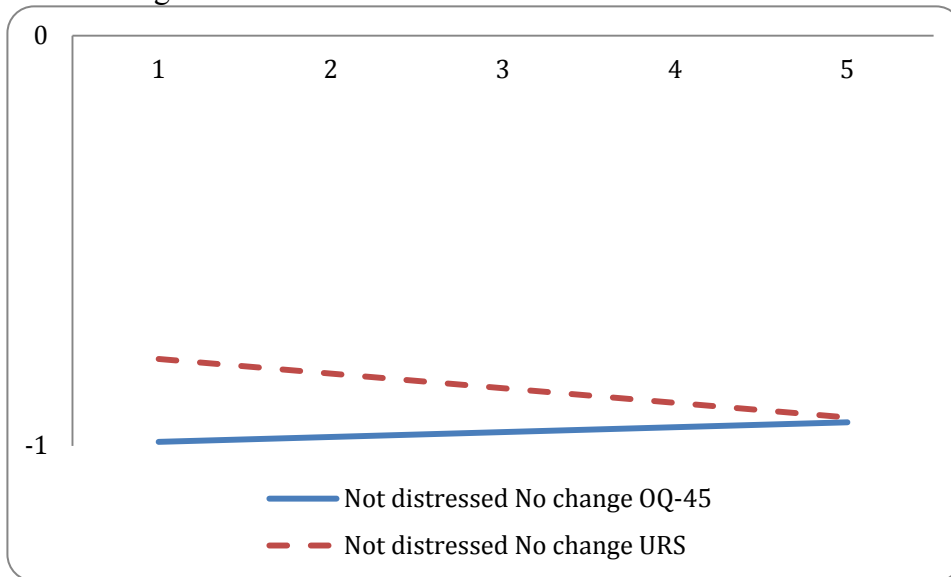
	Estimate	Standard Error	DF	T value	pr>t
Interaction	-0.2726	0.05868	831	-4.65	0.001

Consumers categorized by the OQ-45 as starting treatment in the not distressed range and then improved over time.



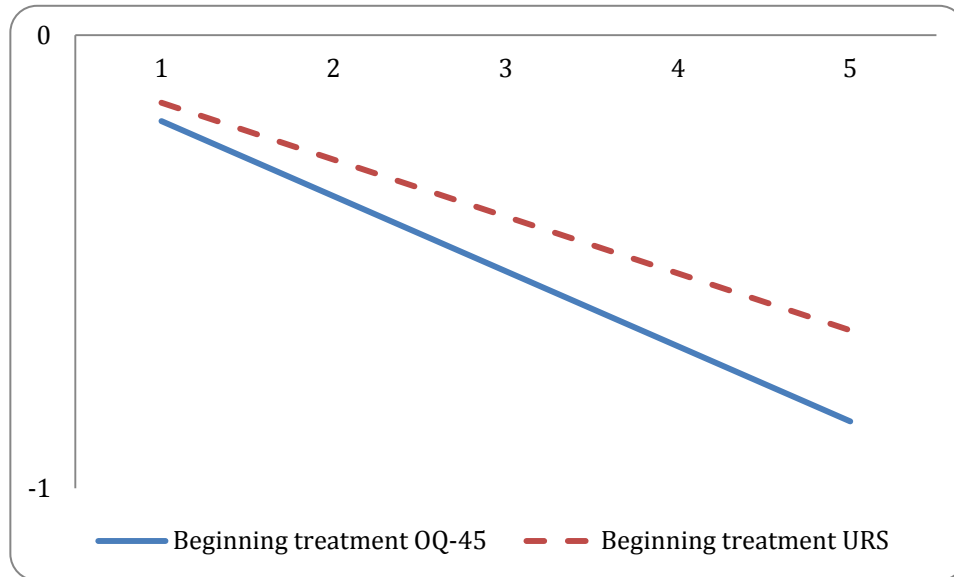
	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.1672	0.06052	103	2.76	0.0068

Consumers categorized by the OQ-45 as starting treatment in the not distressed range and then did not change over time.



	Estimate	Standard Error	DF	T value	pr>t
Interaction	-0.04758	0.02743	300	-1.73	0.0838

Consumers beginning their first episode of treatment.



	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.04011	0.02619	2386	1.53	0.1258

Appendix E: The SAMHSA Fundamental Components of Recovery

Recovery Components	Definition
Self-Direction	Consumers lead, control, exercise choice over, and determine their own path of recovery by optimizing autonomy, independence, and control of resources to achieve a self-determined life. By definition, the recovery process must be self-directed by the individual, who defines his or her own life goals and designs a unique path towards those goals.
Individualized and Person-Centered	Consumers lead, control, exercise choice over, and determine their own path of recovery by optimizing autonomy, independence, and control of resources to achieve a self-determined life. By definition, the recovery process must be self-directed by the individual, who defines his or her own life goals and designs a unique path towards those goals.
Empowerment	Consumers have the authority to choose from a range of options and to participate in all decisions—including the allocation of resources—that will affect their lives, and are educated and supported in so doing. They have the ability to join with other consumers to collectively and effectively speak for themselves about their needs, wants, desires, and aspirations. Through empowerment, an individual gains control of his or her own destiny and influences the organizational and societal structures in his or her life.
Responsibility	Consumers have a personal responsibility for their own self-care and journeys of recovery. Taking steps towards their goals may require great courage. Consumers must strive to understand and give meaning to their experiences and identify coping strategies and healing processes to promote their own wellness.
Strengths-Based	Recovery focuses on valuing and building on the multiple capacities, resiliencies, talents, coping abilities, and inherent worth of individuals. By building on these strengths, consumers leave stymied life roles behind and engage in new life roles (e.g., partner, caregiver, friend, student, employee). The process of recovery moves forward through interaction with others in supportive, trust-based relationships.
Hope	Recovery provides the essential and motivating message of a better future— that people can and do overcome the barriers and obstacles that confront them. Hope is internalized; but can be fostered by peers, families, friends, providers, and others. Hope is the catalyst of the recovery process. Mental health recovery not only benefits individuals with mental health disabilities by focusing on their abilities to live, work, learn, and fully participate in our society, but also enriches the texture of American community life. America reaps the benefits of the contributions individuals with mental disabilities can make, ultimately becoming a stronger and healthier Nation.
Peer Support	Mutual support—including the sharing of experiential knowledge and skills and social learning—plays an invaluable role in recovery. Consumers encourage and engage other consumers in recovery and provide each other with a sense of belonging, supportive relationships, valued roles, and community.
Respect	Community, systems, and societal acceptance and appreciation of consumers —including protecting their rights and eliminating discrimination and stigma—are crucial in achieving recovery. Self-acceptance and regaining belief in one’s self are particularly vital. Respect ensures the inclusion and full participation of consumers in all aspects of their lives.

Recovery Components	Definition
Holistic	Recovery encompasses an individual's whole life, including mind, body, spirit, and community. Recovery embraces all aspects of life, including housing, employment, education, mental health and healthcare treatment and services, complementary and naturalistic services, addictions treatment, spirituality, creativity, social networks, community participation, and family supports as determined by the person. Families, providers, organizations, systems, communities, and society play crucial roles in creating and maintaining meaningful opportunities for consumer access to these supports.
Non-Linear	Recovery is not a step-by-step process but one based on continual growth, occasional setbacks, and learning from experience. Recovery begins with an initial stage of awareness in which a person recognizes that positive change is possible. This awareness enables the consumer to move on to fully engage in the work of recovery.

Appendix F: URS Telephone Survey Script

Instructions: Read all of the instructions in quotation marks verbatim aloud to the participants, and record their answers in the spaces provided.

“Before I start, are you 18 or older” yes/no (circle one)

If the respondent responds no, thank them for their time, and inform them that we are only able to survey individuals above the age of 18. Read them the final statement of the informed consent form and proceed to the next call.

“I am going to start by asking you a couple of demographic questions.”

1. “What is your current age?” _____
2. “What is your gender?” Male/female (circle one)
3. “What is your ethnicity? _____
4. “Are you currently receiving therapeutic or pharmacological treatment for mental illness?”

“Looking back over the last week, including today, please let us know how you have been feeling in relation to these questions over the last week, including today. Please rank the following statements on a scale of 0 to 4 with 0 representing never, 1 representing rarely, 2 representing sometimes, 3 representing frequently, and 4 representing almost always.”

Proceed to read aloud each question followed by reading each possible response allowed (e.g. “I have hope for the future, 0 – never, 1 – rarely, 2 – sometimes, 3 – frequently, 4 – almost always).

1. I have hope for the future.

0	1	2	3	4
Never	Rarely	Sometimes	Frequently	Almost Always

2. I have meaningful work, volunteer, work or activities in my life.

0	1	2	3	4
Never	Rarely	Sometimes	Frequently	Almost Always

3. Most days I get to do something that I enjoy.

0	1	2	3	4
Never	Rarely	Sometimes	Frequently	Almost Always

4. I have a place to live and it’s ok.

0	1	2	3	4
Never	Rarely	Sometimes	Frequently	Almost Always

5. My life has meaning.

0	1	2	3	4
Never	Rarely	Sometimes	Frequently	Almost Always

6. I have people/friends that I can turn to.

0	1	2	3	4
Never	Rarely	Sometimes	Frequently	Almost Always

7. I am connected to my community.

0	1	2	3	4
Never	Rarely	Sometimes	Frequently	Almost Always

8. I am in charge of my own life and recovery.

0	1	2	3	4
Never	Rarely	Sometimes	Frequently	Almost Always

9. I have goals for my future.

0	1	2	3	4
Never	Rarely	Sometimes	Frequently	Almost Always

10. I have meaningful relationships.

0	1	2	3	4
Never	Rarely	Sometimes	Frequently	Almost Always

Upon completion of the last question say the following:

“Thank you for participating in our research survey today. Would you be willing to fill this survey out one more time in 2 to 4 weeks online?”

If the participant says yes, say the following:

“May we have an email address that we can use to send you a link to the survey?”

If the participant says no, proceed with the closing statement from the informed consent script.

Appendix G: URS Total Score Correlations with Individual OQ-45 Items.

OQ-45 Items	r	n
1. I get along well with others.	-.550	8477
2. I tire quickly.	-.425	8487
3. I feel no interest in things.	-.612	8474
4. I feel stressed at work/school	-.332	8425
5. I blame myself for things.	-.488	8490
6. I feel irritated.	-.491	8486
7. I feel unhappy in my marriage/significant relationship.	-.397	8372
8. I have thoughts of ending my life.	-.506	8485
9. I feel weak.	-.504	8484
10. I feel fearful.	-.510	8472
11. After heavy drinking, I need a drink the next morning to get going.	-.066	8452
12. I find my work/school satisfying.	-.594	8374
13. I am a happy person.	-.709	8483
14. I work/study too much.	.096	8442
15. I feel worthless.	-.654	8477
16. I am concerned about family troubles.	-.345	8486
17. I have an unfulfilling sex life.	-.383	8384
18. I feel lonely.	-.582	8488
19. I have frequent arguments.	-.377	8484
20. I feel loved and wanted.	-.702	8491

21. I enjoy my spare time.	-.683	8483
22. I have difficulty concentrating.	-.489	8487
23. I feel hopeless about the future.	-.669	8488
24. I like myself.	-.706	8487
25. Disturbing thoughts come into my mind that I cannot get rid of.	-.475	8486
26. I feel annoyed by people who criticize my drinking (or drug use).	-.094	8460
27. I have an upset stomach.	-.372	8325
28. I am not working/studying as well as I used to.	-.370	8460
29. My heart pounds too much.	-.428	8485
30. I have trouble getting along with friends and close acquaintances.	-.495	8492
31. I am satisfied with my life.	-.744	8501
32. I have trouble at work/school because of drinking or drug use.	-.065	8460
33. I feel that something bad is going to happen.	-.539	8380
34. I have sore muscles.	-.355	8477
35. I feel afraid of open spaces, of driving, or being on buses, subways, and so forth.	-.408	8483
36. I feel nervous.	-.486	8455
37. I feel my love relationships are full and complete.	-.595	8456
38. I feel that I am not doing well at work/school.	-.276	8433
39. I have too many disagreements at work/school.	-.286	8434
40. I feel something is wrong with my mind.	-.537	8483
41. I have trouble falling asleep or staying asleep.	-.396	8490

42. I feel blue.	-.608	8495
43. I am satisfied with my relationships with others.	-.892	8486
44. I feel angry enough at work/school to do something I might regret.	-.339	8453
45. I have headaches.	-.329	8489

Appendix H: Journal Article

Running Head: MEASURING GROWTH

Measuring Growth: The Reliability and Validity of the Utah Recovery Scale

R. Jason Katzenbach, Robert Gleave, Gary Burlingame, & Dallas Jensen

Brigham Young University

Abstract

Objective: An approach to recovery that focuses on helping consumers foster meaningful relationships, independent living, and fulfilling work has been emphasized in the mental health care community. In accordance, the Utah division of the National Alliance on Mental Illness generated recovery indicators (URS) to track consumer treatment progress. This study explored reliability and validity estimates for the URS and the relationship between distress and recovery.

Methods: Community URS data were used in conjunction with archival data from community mental health centers. A clinical cutoff score and reliable change index for the URS were calculated. In addition, URS data were compared to Outcome Questionnaire 45 data using hierarchical linear modeling (HLM).

Results: Evidence supporting the reliability of the URS was found: $\alpha = 0.898$ ($n = 91$). The URS yielded a cutoff score of 31 and a reliable change index of 8. Analyses also indicated a strong inverse relationship between the URS and the OQ-45 ($r = -.75$, $n = 8,483$). In addition, HLM analyses found that the URS tracks consumer progress in a similarly to the OQ-45 ($t = .09$, $df = 9080$, $p = .9282$, $n = 1954$). Subsequent HLM analyses found the URS to be less sensitive than the OQ-45 to consumer changes.

Conclusions: The evidence supports the reliability and validity of the URS's ability to track recovery. The results support the use of the URS as an occasional measure to track consumer treatment response and suggest a strong relationship between recovery and reduction in consumer distress.

Measuring Growth: Reliability and Validity of the Utah Recovery Scale

In 2005, the Substance Abuse and Mental Health Services Administration (SAMHSA) introduced an agenda to change mental health care delivery in the United States. The agenda includes strategies for improving mental health services by implementing evidence-based practices, focusing on recovery, and making treatment consumer/family driven (SAMHSA, 2005). Central to these strategies is the idea that recovery from mental illness is not only possible, but also the expected outcome of effective treatment.

SAMHSA's agenda focuses on a broader definition of recovery that is not limited to symptom amelioration. According to this newer definition, recovery happens when mental health consumers are able to function in society through obtaining meaningful work, fulfilling relationships and independent living (Davidson, et al., 2005; Jacobson, 2004). Recovery in this sense also involves the acquisition of positive coping skills, the restoration of a sense of self, and the pursuit of purposeful living (O'Connor & Delaney, 2007). In addition, recovery from this perspective is thought to relate not only to a reduction in distress but also to an increase in well-being as it has been explored in the positive psychology literature (Anthony 2003; Resnick & Rosenheck 2006). From this perspective, the progress in recovery is assessed collaboratively by the consumer, the care provider, and potentially the consumer's family (Frese, Stanley, Kress, & Vogel-Scibilia, 2001; Jacobson, 2004).

This broader approach to recovery has garnered a large amount of consumer support and has begun to affect mental health treatment policy. Despite the ability of this movement to effect change, very little testing has been done as to whether the implementation of this recovery paradigm is beneficial for consumers. The research that has been done has typically relied strictly on the input of a few consumers in order to shape how recovery is measured, rather than

incorporating a clear definition of recovery into measure creation. Furthermore, no recovery measures have been examined based on their ability to track consumer changes over time.

In order to more clearly define recovery SAMHSA organized a panel of mental health care consumers, family members, providers, advocates, researchers, academicians, managed care representatives, and accreditation organization representatives which together constructed a consensus definition of the recovery model. The resulting consensus statement comprises 10 components that define recovery (SAMHSA, 2006). SAMHSA's components specify a direction for mental health service providers to steer towards in terms of recovery.

In an effort to measure SAMHSA's 10 fundamental recovery components, the Utah division of the National Alliance on Mentally Illness (NAMI-Utah) conducted consumer focus groups in order to develop questions that could assess recovery based on SAMHSA's model. The focus groups developed 10 questions to serve as indicators of recovery. The resulting 10 recovery indicators were labeled the Utah Recovery Scale (URS).

This study examined the psychometric validity of the URS. In addition it also explored the relationship between recovery as measured by the URS and reported consumer distress levels as well as reported quality of life.

Methods

It was necessary to examine the URS in two phases. The first phase of the study was to collect community data for the URS in order to obtain URS data from a non-clinical sample. The next phase of the study examined archival community mental health URS and Outcome Questionnaire 45 (OQ-45) data. Institutional review board approval was obtained from the Brigham Young University IRB for both phases of the study.

Phase 1

Phase 1 of this study was to collect community data from individuals that were not receiving mental health treatment in order to establish a comparison sample to contrast with individuals receiving treatment. These data were collected in October 2009.

Participants. Participants were selected randomly from a local phone book. Trained research assistants contacted participants via telephone and administered the survey according to an administration script. A total of 91 participants were sampled including 53 women and 38 men. Participant ages ranged from 18 to 86 with the average age being 44. Data were not collected for individuals under the age of 18. Demographically, 84 participants identified themselves as Caucasian, 2 as African American, 2 as Asian, 1 as Polynesian, 1 as Latino, and 1 as Mixed Race. After complete description of the study to the participants, informed consent was obtained. Sampled participants that indicated that they were currently receiving some form of mental health treatment were excluded from this study.

Instruments.

The Utah Recovery Scale (URS). The Utah Recovery Scale (URS) was derived from a series of consumer focus groups conducted by the Utah division of the National Alliance on Mentally Illness (NAMI-Utah) in order to construct questions based on SAMHSA's fundamental components of recovery. The focus groups resulted in 10 questions based on SAMHSA's recovery components (SAMHSA, 2005). The URS is scored using a 5-point scale (0= never 1 = rarely, 2 = sometimes, 3= frequently, 4 = almost always), which yields a possible range of scores from 0 to 50.

The URS has no developed reliability or validity statistics. Initial criterion validity for the measure, however, was investigated using additional consumer focus groups that pilot tested the

measure. Participants for three focus groups were selected from three community mental health centers. Two of the focus groups consisted of consumers receiving inpatient treatment while one focus group consisted of participants receiving outpatient treatment. Focus groups ranged from 5 to 15 participants, comprising of a total of 30 participants. The focus group participants, after signing a consent form, were asked to take the URS, and then were asked questions about their experience taking the survey. The focus group leader closely followed a discussion guide of questions and asked follow-up questions when appropriate. The groups lasted between 35 and 50 minutes and the groups were digitally recorded for further analysis. The recordings were reviewed in conjunction with notes from the focus groups and analyzed for emergent themes.

Response to the URS from the focus groups was, for the most part, positive. Almost all of the consumers across the groups said that the questions addressed elements of recovery that were important to them—something that they did not always feel other outcome measures they had been exposed to had done.

Phase 2

Data. Consumer data for this study were taken from the archival records of community mental health centers in Utah. Community mental health centers that receive state funding track treatment progress using a variety of measurement tools including the Outcome Questionnaire 45 (OQ-45) and the URS. Consumer data are then databased and reported back to clinicians using the OQ-45 analyst system that was developed and supported by OQ-45 Measures LLC. This database provides outcome data for each consumer. Approximately 30,000 adults receive mental health services from the state of Utah on a yearly basis. Outcome data from inpatient, residential, and outpatient consumers were analyzed.

Instruments. In addition to the URS this phase of the study also examined consumer distress as measured by the OQ-45.

The Outcome Questionnaire-45 (OQ-45). The Outcome Questionnaire-45 (OQ-45) is an instrument designed to measure consumer outcomes in therapeutic settings (Lambert et al., 2004). The OQ-45 is a 45 item self-report questionnaire scored using a 5-point scale (0= never 1 = rarely, 2 = sometimes, 3= frequently, 4 = almost always) that yields a possible range of scores from 0 to 180. High scores on the OQ-45 indicate more distress and as consumers improve scores decrease. The OQ-45 has been validated across cultures using a variety of normal and consumer populations (Lambert et al., 2004). It has become a gold standard for measuring distress as it relates to treatment, and was recently found to be the 3rd most commonly used measure of outcome in a survey by psychologists in clinical practice (Hatfield & Ogles, 2004).

The OQ-45 takes approximately 5-7 minutes to complete and is typically administered prior to each treatment session. Concurrent validity is moderate to high ($r = 0.50 - 0.85$) when correlated with measures most often used to assess psychotherapy outcome in clinical trials (Lambert et al., 2004). Most importantly, the OQ-45 has been shown to be sensitive to changes in consumers over short time periods, while remaining stable in untreated individuals (Vermeersch, Lambert, & Burlingame, 2000 & Vermeersch et al., 2004). The OQ-45 has a reported 3-week test-retest reliability value of $r = .84$ and a reported internal consistency of $r = .93$ (Lambert et al., 2004). Reliable change indices (RCI) for the OQ-45 have been calculated using formulas developed by Jacobson and Truax (1991). The RCI for the OQ-45 was calculated to be 14 points using normative data from community non-consumers ($N = 1353$) and consumers entering treatment ($n = 1476$), thus consumers who exhibit a 14 point positive or negative change are found to have made reliable change. A clinical cutoff score of the OQ-45 was found to be

63/64, thus when a consumer's score drops under 64 they are thought to be functioning more like non-consumers than typical consumer populations. When a consumer's score has dropped by 14 points or more, and fallen under 64, then the consumer has met the criteria for clinically significant change (Ellsworth et al., 2006).

In addition to the total score the OQ-45 has three subscales that measure quality of interpersonal relations, social role functioning and symptom distress (Lambert, et al., 2004). The Subjective Distress subscale is a 25-item scale that evaluates symptoms such as depression and anxiety. The Interpersonal Relationship subscale consists of 11 items that attempt to assess functioning in interpersonal relationships. The final subscale, Social Role consists of 9 items that attempts to measure dysfunction in roles such as work and leisure life activities. The OQ-45 also contains several reverse scored quality of life items designed to measure things like quality of relationships and overall life satisfaction

Statistical Analysis

Hierarchical linear modeling (HLM) was used to examine the longitudinal relationship between the URS and the OQ-45. Research examining treatment trajectory growth curves is ideally suited to the use of HLM that can be generated from consumer responses to outcome measure items (Arnold, 1992). In addition, HLM accounts for missing data so that missing participant data does not need to be thrown out because of the limitations of the model (Speer & Greenbaum, 1995). The HLM analyses for this study were conducted using PROC MIXED in SAS.

Results

To explore the reliability of the URS participants were randomly selected from the community. After giving consent participants answered a few demographic questions and then

the URS items over the phone. Six participants that indicated that they were currently receiving some form of mental health treatment were excluded from this analysis. The overall URS internal consistency estimate was high ($\alpha = 0.898$, $n = 91$).

Next, archival consumer data from community mental health centers in the state of Utah were used in order to explore the relationship between recovery and consumer distress using the OQ-45. Initial analyses indicated a strong inverse relationship between the two measures ($r = -0.75$, $n = 8,483$). The strong inverse relationship indicates that consumers that scored higher on recovery as measured by the URS scored lower on distress as measured by the OQ-45. In addition, several individual items from the two measures had moderately strong negative correlations ($r < -.6$) and are reported in Table 1. In addition, moderate to strong relationships were also found between the URS and each of the OQ-45 subscales: Symptom Distress ($r = -.711$, $n = 7878$), Social Role Functioning ($r = -.576$, $n = 8,197$), and Interpersonal Relations ($r = -.746$, $n = 8162$).

The OQ-45 contains several reverse scored quality of life items designed to measure things like quality of relationships and overall life satisfaction. As with the rest of the OQ-45, higher scores on these items are indicative of higher levels of distress. A composite score was created from these items and then compared with the URS total score. A strong inverse relationship was shown between the quality of life items (higher scores indicate higher levels of distress) and the URS (higher scores indicate higher levels of recovery) ($r = -.863$, $n = 8,216$) (consumers with missing scores on these particular items were excluded from this analysis).

Next, archival longitudinal consumer data from community mental health centers in Utah were explored in order to assess the URS's ability to detect changes among consumers. This sample consisted of 1445 consumers that took the URS over several administrations ranging

from 2 to 16 with the average number of administrations being 3.5. The first step in this process was to calculate a cutoff point, or the point along the range of possible URS scores that best separates the distribution of the non-consumer scores from the distribution of consumer scores. The cutoff point between adjacent samples defines the point where it is statistically more likely for a score to be in one, as opposed to the adjacent overlapping distribution (Jacobson et al., 1984). The calculation is as follows:

$$c = \frac{(SD_1)(Mean_2) + (SD_2)(Mean_1)}{SD_1 + SD_2}$$

Where SD_1 is the standard deviation of the clinical (consumer) population, $Mean_2$ is the mean of the non-consumer population, SD_2 is the standard deviation of the non-consumer population, and $Mean_1$ is the mean of the clinical (consumer) population. In order ensure homogeneity of the consumer sample, consumers whose OQ-45 score fell below the clinical cutoff of 63 were excluded from the analysis. Descriptive data for these two samples on the URS are reported in Table 2.

Prior to calculating a cutoff point it was first necessary to ensure that both samples used in the calculation were statistically distinct (Tingey et al., 1996). Distinctness of the samples was determined using an independent samples t test and a “ d ” test. The samples met both criteria for statistical distinctness. The independent samples t value of -27.841 ($p < .05$, $df = 95.577$) surpassed alpha of .05. In addition, the calculated “ d ” value surpassed the criterion of .5 further indicating the distinctiveness of the two samples ($d = 2.56$). Once distinct samples were statistically verified a cutoff point of 28.799 (28/29) was calculated.

The cutoff score enables the URS to function in distinguishing consumers from non-consumers. Table 3 presents the accuracy with which the cutoff score classified respondents. The sensitivity of .89 indicates that of actual consumers, 89% scored below the clinical cutoff score

of 29 and 11% scored above. In other words, use of the cutoff score identified 89% of actual consumers. The specificity of .95 indicates that of the actual non-consumers, 95% scored above the cutoff score and 5% scored below. Thus the cutoff score identified 95% of actual non-consumers. The overall accuracy of the cutoff score in predicting consumer versus non-consumer status was 0.89 (hit rate). The positive predictive power (PPP) of .99 indicates that of the cases the cutoff score predicted to be consumers, 99% of them were actual consumers. The negative predictive power (NPP) of .14 indicates that of the cases the cutoff score predicted to be non-consumers, 14% of them were actual non-consumers. Note that whereas the test characteristics of sensitivity, specificity, and hit rate are relatively independent of prevalence of the condition being tested, the characteristics of PPP and NPP are highly dependent upon prevalence (Streiner, 2003). Thus the high PPP and low NPP are expected considering the large proportion of consumers (98%) versus non-consumers (2%) in the sample producing these indices.

In addition to a cutoff score, a reliable change index (RCI) was also calculated for the URS. The RCI is a way of determining whether changes observed on a measure are greater than what would be attributed to measurement error (Tingey et al., 1996). The RCI is calculated by multiplying the standard error of the difference by the point location on a distribution in order to achieve a certain confidence interval. The RCI for the URS was calculated at the 95% confidence level. In addition, the reliability coefficient used in the equation was Cronbach's alpha coefficient of .898 obtained from the community sample. The equation for calculating the RCI is presented below:

$$95\%RCI = S_{diff} = (4.095)(1.96) = 8.062 = 8/9$$

$$S_{diff} = \sqrt{2(S_E)^2} = \sqrt{2(2.896)^2} = 4.095$$

$$S_E = SD_p \sqrt{1 - \alpha} = 9.066 \sqrt{1 - 0.898} = 2.896$$

$$SD_p = \sqrt{\frac{(n_1 - 1)(SD_1^2) + (n_2 - 1)(SD_2^2)}{(n_1 + n_2 - 2)}} = \sqrt{\frac{(1542 - 1)(8.983) + (1445 - 1)(9.1554)}{(1542 + 1445 - 2)}} = 9.066$$

where 95% RCI is the RCI value at the 95% confidence level, S_{diff} is the standard error of the difference, S_E is the standard error of measurement, SD_p is the pooled standard deviation for the first and last score used in the calculation, α is the internal consistency reliability utilized for the calculation, n_1 is the sample size for the first administration, n_2 is the sample size for the last administration, SD_1 is the standard deviation of the first administration of the URS, and SD_2 is the Standard deviation of the last administration of the URS. The calculated RCI for the URS was 8.026 indicating that consumers who have changes greater than 8 points in their total URS score have made clinically significant change.

The RCI and the cutoff score were used to determine consumer improvement, deterioration, and stability. In addition, consumers that reached the clinical cutoff of score of greater than 28 were considered recovered. Consumers whose first and last scores did not have a corresponding OQ-45 score because of missing data were excluded from this analysis. The results are presented in the table 5. Next consumers were compared based on their change status as calculated by the URS with their change status as calculated by the OQ-45. The results are presented in Table 5.

In order to investigate the URS's sensitivity to consumer changes over time the consumers' total scores and individual item scores were compared against zero (or no change). HLM was used for this analysis because there are multiple waves of data among consumers with differing numbers of administrations per consumer. Results of the initial data analysis indicated that the URS total score and all 10 items met the first criterion for change sensitivity in that they

demonstrated change in the theoretically proposed direction (i.e., consumers improved over time as illustrated by a positive slope). Of these 10 items, 7 demonstrated a slope that was significantly different from zero. The results are presented in Table 6.

Next, treatment trajectories measured by the URS were compared to treatment trajectories measured by the OQ-45 using HLM. Prior to conducting these statistical analyses, reverse scoring procedures were performed on URS data so that lower scores indicated improvement similarly to the OQ-45. In addition, scores from the two tests were standardized using *Z* scores so that both measures were on an equal scale. Results of the initial data analysis indicated that the URS tracked change in the theoretically proposed direction (i.e., consumers improved over time at a similar rate as measured by the OQ-45). In groups in which the slopes of the two measures were statistically different the OQ-45 had a steeper slope suggesting that the URS may not be as sensitive to certain types of changes in consumers. The results are presented in Table 7.

Discussion

This study examined the psychometric validity of the URS. The URS was examined in several psychometric categories including reliability, validity, sensitivity to change, and validity for change. First, data were collected from a randomly sampled community population in order to explore the reliability of the URS. Next, archival URS data collected from mental health consumers in the state of Utah were also examined in order to answer questions regarding the validity of the measure as well as the measure's ability to detect clinically meaningful changes among consumers.

Prior to examining the psychometric validity of the URS, its reliability was first examined. The obtained internal consistency estimate of $\alpha = 0.898$ ($n = 91$) provides initial evidence for the reliability of the URS.

Next, the criterion related validity of the URS was explored. Theoretically, a high score on a recovery measure should have a moderate negative correlation with consumer distress. Initial analyses indicate a strong inverse relationship between the two measures ($r = -0.75$, $n = 8,483$). This strong relationship provides evidence that consumers with higher recovery scores are likely to have lower distress scores and vice versa. This finding provides some evidence for the validity of the URS because recovery and distress are thought to be inversely related. This conclusion is also supported by the URS's strong relationship to the Symptom Distress subscale of the OQ-45 ($r = -.711$, $n = 7878$). This stronger than expected relationship may also provide evidence that the constructs of recovery and consumer distress actually overlap to an extent that measuring recovery separately may be unnecessary. In addition, this finding provides initial evidence that consumers receiving recovery-focused treatments are more likely to show improvement in reduction of distress. Since this relationship is correlational more research is needed in order to further explore this possibility. In addition, the URS's demonstrated strong relationship to the OQ-45 subscale of Interpersonal Relations ($r = -.746$, $n = 8162$) and moderate relationship to the OQ-45 subscale of Social Role ($r = -.576$, $n = 8197$) provide further evidence for the construct validity for the URS by demonstrating that consumers exhibiting higher levels of recovery exhibit lower levels of relationship problems, and lower levels of role dissatisfaction.

In addition, a strong inverse relationship was shown between the reverse scored quality of life items (higher scores indicate higher levels of distress) and the URS (higher scores indicate higher levels of recovery) ($r = -.863$, $n = 8,216$). This strong relationship indicates that recovery

seems to be highly related to consumer quality of life as expected. This strong relationship provides initial evidence for the possible link between recovery and factors that may contribute to well being.

Sensitivity and Specificity

Next, the URS's ability to differentiate between clinical and non-clinical populations was examined. The URS was generally accurate in both its identification of actual consumers (sensitivity = .89) as well as its identification of actual non-consumers (specificity = .95). The ability of the URS to identify both consumers and non-consumers accurately suggests its cutoff score could be useful for identifying when consumers move from a clinical population to a non-clinical population during treatment

Next, the URS's sensitivity to changes among consumers in treatment was explored. In order to investigate the URS's ability to detect change the URS was compared against 0 or no change. The URS total score was demonstrated to be significantly different from zero when tracking consumers in treatment over time. The observed changes also occurred in the theoretically specified direction (i.e., consumers in treatment improved). In terms of the individual items, all 10 of items demonstrated change in the theoretically specified direction; however, only 7 of the ten items demonstrated a rate of change that was statistically significant when compared to zero. This finding provides evidence that the URS as a whole as well as 7 of the 10 items are able to detect changes in consumers in treatment over time. The three items that did not detect significant change when compared to zero were "I have a place to live and it's ok,"; "I have people/friends I can turn to,"; "I have meaningful relationships." There are a number of possibilities that may have contributed to the inability of these items to detect changes. One possibility is that these questions are poor indicators of aspects of consumers'

lives that are likely to change during treatment. Another possibility is that outside relationships and living situations are not emphasized enough in the mental health treatment programs in Utah for there to be measured changes in those areas. A third possibility is that the areas that these items measure may change more slowly over time than other areas measured by the URS. Yet another possibility could be that these items cover areas that may change once but not gradually over time (i.e., once someone has a place to live that they are ok with they are not likely to have this progress any more). One potential solution to improve the sensitivity to change of the URS would be to remove these items; however, if the inability of these items to detect changes that are significantly different from no change is due to a lack of emphasis in these recovery areas in consumer treatment removing these items would only serve to further take attention away from these recovery areas deemed important by consumers. Further inquiry is needed to ascertain the cause of the failure of these items to detect changes.

Validity for Change

In addition to sensitivity to change validity for change was also explored. When HLM was used to compare consumer treatment trajectories between the two measures the two measures were found to track change in a similar way. When the two measures were examined among different classifications of consumers the measures were again found to be mostly similar. In cases where the slopes of the two measures were significantly different there was no difference in the directionality of the slopes (positive vs negative) and consistently the URS had a shallower slope. These observed differences may suggest that the two domains are somewhat different in terms of the way consumers change throughout treatment. One explanation for the observed differences between the measures could be that consumer changes in recovery occur more gradually than consumer changes in distress. For this reason it may be more useful to

administer recovery instruments like the URS less frequently than distress instruments like the OQ-45. This approach may also be practically beneficial so-as not to overly burden consumers with the frequent survey administrations.

In addition, the URS's ability to detect reliable changes in consumers over time (Jacobson, Truax, & Kazdin, 1992). The ability of the URS to detect reliable changes was examined by calculating an RCI for the measure and then comparing this with the established RCI of the OQ-45. Utilizing the RCI the URS identified 12% of consumers as recovered (scoring within a range that is most likely to be in a non-clinical population), 77% of consumers as showing no change, 5% of consumers as improved (moving significantly towards scores that are likely to fall in a non-clinical population), and 6% of consumers as deteriorated (moving significantly away from scores that are likely to fall within a non-clinical population).

When compared, the URS identified the proportion of consumers in the same category as the OQ-45 58% of the time. However, when broken down into individual groups the URS change categorizations were equal to the OQ-45 to a much lesser extent with the URS identifying only 42% of the consumers classified as recovered by the OQ-45, 12% of the consumers classified as improved by the OQ-45, and 23% of the consumers classified as deteriorated by the OQ-45. In all cases, with the exception of the no change classification, the URS identified fewer consumers in each category than the OQ-45. This fits with the finding that recovery as measured by the URS may be less sensitive to clinical changes than distress as measured by the OQ-45. For this reason, the URS may be most useful as an occasional outcome measure

Limitations

One of the major remaining questions regarding the validity of the URS is test-retest reliability. A high test-retest reliability estimate observed in a group of consumers not receiving treatment would provide evidence that the URS is actually detecting changes that are the result of treatment. Further, multiple administrations given to a group of consumers not receiving treatment could also be compared against consumers receiving treatment to explore whether the URS as a measure as well as its individual items are sensitive to change over time. Comparing the URS to a sample of individuals not in treatment would provide further evidence for the sensitivity of the URS for change because consumers not receiving treatment may show improvement over time (Aneshensel, Estrada, Hansell, & Clark, 1987; Bromet, Dunn, Connell, Dew, & Schulberg, 1986; Durham, Burlingame, & Lambert, 1998; Henderson, Byrne, & Duncan-Jones, 1981; Jorm, Duncan-Jones, & Scott, 1989).

Additionally, consumers were not given the URS at intake. Because of this we were unable to examine consumer treatment trajectories that started at intake. We were able to look at consumers beginning treatment in comparison to the OQ-45 but most often these consumers were not given the URS until their second or third session. Examining a group of consumers that take the URS at intake and then tracking them throughout treatment might give us not only a more clear picture of recovery's longitudinal relationship to distress but also might allow us to examine time differences between the two (e.g., perhaps a certain level of distress reduction is required before change in recovery can be observed).

Lastly, it is important to note that the URS was developed utilizing consumer feedback and focus groups and not from a large pool of items. This is important because it is possible that utilizing a large pool of items in its construction would have provided a more conservative

approach to insuring that the domain of recovery was adequately captured. For this reason, further exploration is needed in comparing the URS with other measures of recovery that are thought to cover multiple domains of recovery such as the Recovery Assessment Scale (Corrigan et al., 1999).

Conclusions

Initial evidence for the reliability, validity, sensitivity for change, and validity for change suggest that the URS may be a useful instrument for clinicians and community mental health programs that are seeking to measure recovery among consumers over time. Given the results of the study, the URS appears to be most useful as an occasional evaluative measure of recovery.

The strong inverse relationship observed between recovery and distress is encouraging for community mental health centers implementing recovery-oriented programs because it indicates that recovery-oriented interventions are unlikely to detract from and may enhance treatment efforts that are aimed at reducing consumer symptoms or distress. Further the strength of the relationship between the two measures may even suggest that recovery-oriented interventions bolster consumer gains in distress reduction. Since this study was not experimental further inquiry is needed in order to fully explore the nature of this finding.

In addition, the strong relationship between the OQ-45 quality of life items and the URS also provide evidence for the potential relationship between recovery and well being. For this reason, it might be helpful for clinicians to look towards positive psychology research, particularly in areas of engagement, reported well being, and social relationships as a way of developing and implementing recovery oriented interventions.

References

- Andresen, R., Caputi, P., & Oades, L. (2006). Stages of Recovery Instrument: Development of a measure of recovery from serious mental illness. *Australian and New Zealand Journal of Psychiatry, 40*(11), 972-980.
- Aneshensel, C. S., Estrada, A. L., Hansell, M. J., & Clark, V. A. (1987). Social psychological aspects of reporting behavior: Lifetime depressive episode reports. *Journal of Health and Social Behavior, 28*, 232–246.
- Anthony, W. A. (2003). Expanding the evidence base in an era of recovery. *Psychiatric Rehabilitation Journal, 27*(1), 1-2.
- Arnold, C. (1992). An introduction to hierarchical linear models. *Measurement & Evaluation in Counseling & Development, 25*(2), 58.
- Bromet, E. J., Dunn, L. O., Connell, M. M., Dew, M. A., & Schulberg, H. C. (1986). Long-term reliability of diagnosing lifetime major depression in a community sample. *Archives of General Psychiatry, 43*, 435–440.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*(1), 147-158.
- Burlingame, G. M., Lambert, M. J., Reisinger, C. W., & Neff, W. M. (1995). Pragmatics of tracking mental health outcomes in a managed care setting. *Journal of Mental Health Administration, 22*(3), 226-236.
- Campbell-Orde, T., Chamberlin, J., Carpernter, J., & Leff, H. S. (2005). *Measuring the promise: A Compendium of Recovery Measures* (Vol. 2). Cambridge, MA: The Evaluation Center at the Human Services Research Institute.
- Corrigan, P. W., Giffort, D., Rashid, F., Leary, M., & Okeke, I. (1999). Recovery as a

- psychological construct. *Community Mental Health Journal*, 35(3), 231-239.
- Corrigan, P. W., Salzer, M., Ralph, R. O., Sangster, Y., & Keck, L. (2004). Examining the Factor Structure of the Recovery Assessment Scale. *Schizophrenia Bulletin*, 30(4), 1035-1041.
- Davidson, L., Lawless, M. S., & Leary, F. (2005). Concepts of recovery: Competing or complementary? *Current Opinion in Psychiatry*, 18(6), 664-667.
- Davidson, L., O'Connell, M., Tondora, J., Styron, T., & Kangas, K. (2006). The top ten concerns about recovery encountered in mental health system transformation. *Psychiatric Services (Washington, D.C.)*, 57(5), 640-645.
- Davidson, L., O'Connell, M. J., Tondora, J., Lawless, M., & Evans, A. C. (2005). Recovery in Serious Mental Illness: A New Wine or Just a New Bottle? *Professional Psychology: Research and Practice*, 36(5), 480-487.
- Davidson, L., & Roe, D. (2007). Recovery from versus recovery in serious mental illness: One strategy for lessening confusion plaguing recovery. *Journal of Mental Health*, 16(4), 459-470.
- Durham, C., Burlingame, G. M., & Lambert, M. J. (1998). *Outcome Questionnaire: Repeated administrations, mechanical responding, and social desirability*. Unpublished manuscript, Department of Psychology, Brigham Young University, Provo, UT.
- Ellsworth, J. R., Lambert, M. J., & Johnson, J. (2006). A Comparison of the Outcome Questionnaire-45 and Outcome Questionnaire-30 in Classification and Prediction of Treatment Outcome. *Clinical Psychology & Psychotherapy*, 13(6), 380-391.
- Fisher, D. B., & Ahern, L. (2002). Evidence-based practices and recovery. *Psychiatric Services*, 53(5), 632-633.
- Frese, F. J., III, Stanley, J., Kress, K., & Vogel-Scibilia, S. (2001). Integrating evidence-based

- practices and the recovery model. *Psychiatric Services*, 52(11), 1462-1468.
- Ghiselli, E., Campbell, J., & Zedeck, S. (1981). *Measurement Theory for the Behavioral Sciences*. San Francisco, CA: W.H. Freeman and Company.
- Glaros, Alan G.; Kline, & Rex B. (1988). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. *Journal of Clinical Psychology*, 44(6), 1013-1023.
- Guyatt, G., Walter, S., & Norman, G. (1987). Measuring change over time: assessing the usefulness of evaluative instruments. *Journal Of Chronic Diseases*, 40(2), 171-178.
- Hatfield, D. R., & Ogles, B. M. (2004). The Use of Outcome Measures by Psychologists in Clinical Practice. *Professional Psychology: Research and Practice*, 35(5), 485-491.
- Henderson, A. S., Byrne, D. G., & Duncan-Jones, P. (1981). *Neurosis and the social environment*. New York: Academic.
- Jacobson, N. (2004). *In recovery: The making of mental health policy*. Nashville, Tennessee: Vanderbilt University University Press.
- Jacobson, N., & Curtis, L. (2000). Recovery as policy in mental health services: Strategies emerging from the states. *Psychiatric Rehabilitation Journal*, 23(4), 333-341.
- Jacobson, N., & Greenley, D. (2001). What is recovery? A conceptual model and explication. *Psychiatric Services*, 52(4), 482-485.
- Jacobson, N.S., Follette, W.C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15, 336-452.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical*

Psychology, 59(1), 12-19.

- Jacobson, N. S., Truax, P., & Kazdin, A. E. (1992). Clinical significance : A statistical approach to defining meaningful change in psychotherapy research. In *Methodological issues & strategies in clinical research*. (pp. 631-648). Washington, DC, US: American Psychological Association.
- Jorm, A. F., Duncan-Jones, P., & Scott, R. (1989). An analysis of the re-test artifact in longitudinal studies of psychiatric symptoms and personality. *Psychological Medicine*, 19, 487–493.
- Kazdin, A. E. (1998). *Research design in clinical psychology (3rd ed.)*. Needham Heights, MA US: Allyn & Bacon.
- Keyes, C. L. M. (2007). Promoting and protecting mental health as flourishing: A complementary strategy for improving national mental health. *American Psychologist*, 62(2), 95-108.
- Lambert, M. J., Gregersen, A. T., Burlingame, G. M., & Maruish, M. E. (2004). The Outcome Questionnaire-45. In *The use of psychological testing for treatment planning and outcomes assessment: Volume 3: Instruments for adults (3rd ed.)*. (pp. 191-234). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Lambert, M. J., Hill, C. E., Bergin, A. E., & Garfield, S. L. (1994). Assessing psychotherapy outcomes and processes. In *Handbook of psychotherapy and behavior change (4th ed.)*. (pp. 72-113). Oxford England: John Wiley & Sons.
- Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on consumer progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research*, 11(1), 49-68.

- O'Connor, F. W., & Delaney, K. R. (2007). The recovery movement: Defining evidence-based processes. *Archives of Psychiatric Nursing, 21*(3), 172-175.
- Okiishi, J., Lambert, M. J., Eggett, D., Nielsen, L., Dayton, D. D., & Vermeersch, D. A. (2006). An analysis of therapist treatment effects: toward providing feedback to individual therapists on their consumers' psychotherapy outcome. *Journal of Clinical Psychology, 62*(9), 1157-1172.
- Okiishi, J., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An empirical analysis of therapist effects. *Clinical Psychology & Psychotherapy, 10*(6), 361-373.
- Peysers, H. (2001). Commentary: What is recovery? A commentary. *Psychiatric Services, 52*(4), 486-487.
- Peysers, H. S. (2001). 'A conceptual model of recovery': Reply. *Psychiatric Services, 52*(5), 688-689.
- Raudenbush, S. W., & Chan, W.-S. (1993). Application of a hierarchical linear model to the study of adolescent deviance in an overlapping cohort design. *Journal of Consulting and Clinical Psychology, 61*(6), 941-951.
- Resnick, S. G., & Rosenheck, R. A. (2006). Recovery and positive psychology: Parallel themes and potential synergies. *Psychiatric Services, 57*(1), 120-122.
- Roe, D., & Kravetz, S. (2003). Different Ways of Being Aware of a Psychiatric Disability: A Multifunctional Narrative Approach to Insight into Mental Disorder. *Journal of Nervous and Mental Disease, 191*(7), 417-424.
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57*(6), 1069-

1081.

- Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, 69(4), 719-727.
- Substance Abuse and Mental Health Services Administration (SAMHSA), (2005). *Transforming mental health care in America. The Federal Action agenda: First Steps*. Retrieved September 5, 2007, from http://www.samhsa.gov/Federalactionagenda/NFC_TOC.aspx.
- Substance Abuse and Mental Health Services Administration (SAMHSA), (2006). *National consensus statement on mental health recovery*. Retrieved September 10, 2007, 2007, from <http://mentalhealth.samhsa.gov/publications/allpubs/sma05-4129/>
- Sawyer, K. (2002). Toward a Positive Psychology. *PsycCRITIQUES*, 47(6), 691-693.
- Seligman, M. E. P., Peterson, C., Aspinwall, L. G., & Staudinger, U. M. (2003). Positive clinical psychology. In *A psychology of human strengths: Fundamental questions and future directions for a positive psychology*. (pp. 305-317). Washington, DC, US: American Psychological Association.
- Seligman, M. E. P., Steen, T. A., Nansook, P., & Peterson, C. (2005). Positive psychology progress: Empirical validation of interventions. *American Psychologist*, 60(5), 410-421.
- Speer, D. C., & Greenbaum, P. E. (1995). Five methods for computing significant individual consumer change and improvement rates: Support for an individual growth curve approach. *Journal of Consulting and Clinical Psychology*, 63(6), 1044-1048.
- Tingey, Lambert, Burlingame, & Hansen (1996). Assessing clinical significance: Proposed extensions to method. *Psychotherapy Research*, 6(2), 109-123.
- Vermeersch, D. A. (1998). *Sensitivity to Change as an Item Selection Criteria for an Outcome Measure*. Unpublished Dissertation, Brigham Young University, Provo, UT.

Vermeersch, D. A., Lambert, M. J., & Burlingame, G. M. (2000). Outcome questionnaire: Item sensitivity to change. *Journal of Personality Assessment*, 74(2), 242-261.

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 Consumers? *Journal of Counseling Psychology*, 51(1), 38-49.

Ware, J. H. (1985). Linear models for the analysis of longitudinal statistics. *American Statistician*, 39, 95-101.

Yerushalmy, J. (1947). Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. *Public Health Reports*. 62, 1432-1449.

Table 1

Item Correlations Between the OQ-45 and the URS

URS items	OQ-45 items	r	n
1. I have hope for the future.	13. I am a happy person.	-0.634	8739
	15. I feel worthless.	-0.615	8738
	23. I feel hopeless about the future.	-0.671	8746
	24. I like myself.	-0.643	8742
	31. I am satisfied with my life.	-0.67	8761
	OQ-45 total	-0.67	8750
2. Most days I get to do something that I enjoy.	—	—	—
3. I do something I enjoy during the day.	21. I enjoy my spare time.	-0.65	8726
	31. I am satisfied with my life.	-0.611	8744
	OQ-45 total	-0.619	8734
	4. I feel the place I live is ok.	—	—
5. My life has meaning.	13. I am a happy person.	-0.624	8714
	20. I feel loved and wanted	-0.617	8731
	24. I like myself.	-0.638	8716
	31. I am satisfied with my life.	-0.683	8736
	OQ-45 total	-0.648	8724
6. I have people/friends that I can turn to.	20. I feel loved and wanted	-0.612	8722
7. I am connected to my community.	—	—	—
8. I am in charge of my own life and recovery.	—	—	—
9. I have goals for my future.	—	—	—
10. My relationships are meaningful.	20. I feel loved and wanted	-0.632	8717
	43. I am satisfied with my relationships with others.	-0.617	8705
	3. I feel no interest in things.	-0.612	8474
	13. I am a happy person.	-0.709	8483
URS total	20. I feel loved and wanted	-0.702	8491
	21. I enjoy my spare time.	-0.663	8483
	24. I like myself.	-0.706	8487
	42. I feel blue	-0.608	8495
	43. I am satisfied with my relationships with others.	-0.692	8486

Note. — indicates that correlations for that item were weaker than .6.

Table 2

<i>Descriptive Data for the Community and Clinical Samples</i>			
Sample	Mean	SD	n
Clinical	19.67	7.072	5,010
Community	35.84	5.455	91

Table 3

<i>Sensitivity and Specificity of the URS Clinical Cutoff Score</i>				
		Actual		Total
		Positive (consumer)	Negative (non-consumer)	
Predicted	Positive (consumer)	True Positives 86	False Positives 536	622 12%
	Negative (non-consumer)	False Negatives 5	True Negatives 4474	4479 88%
	Total	91 2%	5010 98%	5101 100%
Sensitivity	Specificity	Hit Rate	Positive Predictive Power	Negative Predictive Power
0.89	0.95	0.89	0.99	0.14

Table 4

<i>URS Change Status</i>		
	n	% of total
Total	1445	100%
Recovered	176	12%
No change	1114	77%
Improved	67	5%
Deteriorated	88	6%

Table 5

Consumer change status classified by the URS and the OQ-45.

	Total	Recovered	No change	Improved	Deteriorated
OQ-45	1445	145	815	251	234
Matching NRI change status	842	61	699	29	53
<i>p</i> of matching status	0.58	0.42	0.86	0.12	0.23
NRI Recovered	—	—	70	38	7
NRI no change	—	65	—	178	172
NRI Improved	—	15	21	—	2
NRI deteriorated	—	4	25	6	—

Table 6

URS Total and Individual Item Slopes Compared to Zero

Item	Slope	t
Total	0.07654	5.87*
1. I have hope for the future.	0.04967	3.18*
2. I have meaningful work/volunteer activities in my life.	0.06763	4.03*
3. Most days I get to do something that I enjoy.	0.07075	4.47*
4. I have a place to live and it's ok.	0.01913	1.16
5. My life has meaning.	0.0747	5.16*
6. I have people/friends I can turn to.	0.02135	1.36
7. I am connected to my community	0.075	4.99*
8. I am in charge of my own life and recovery.	0.07862	4.57*
9. I have goals for the future.	0.05082	3.48*
10. I have meaningful relationships	0.02817	0.078

N = 1954**p* < .05.

Table 7

Slope Comparisons by Change Classification Between the URS and the OQ-45

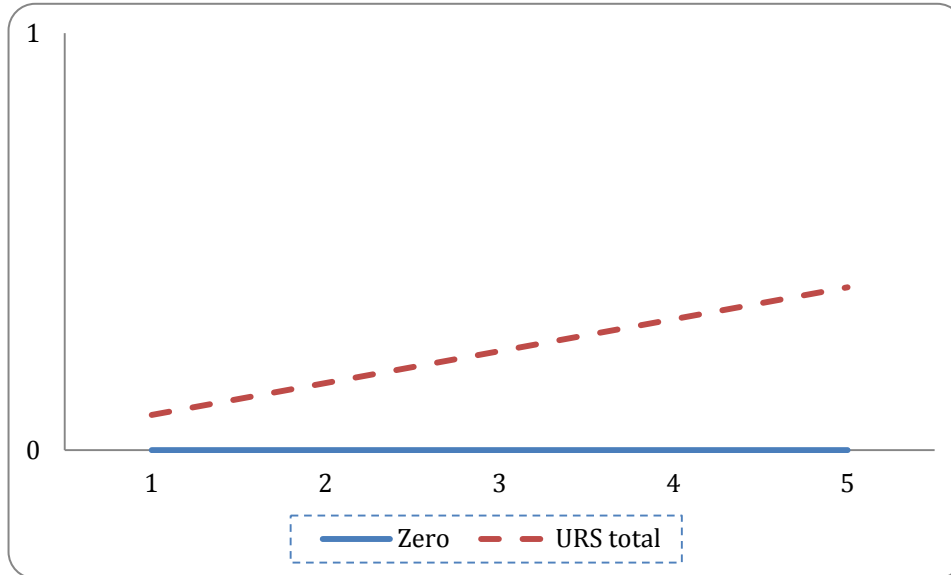
Classification	Slope URS	Slope OQ-45	n	t
Total	-0.079204	-0.08045	1954	0.09
Distressed	-0.14791	-0.1921	1151	2.74*
Distressed improved	-0.3765	-0.5247	203	4.07*
Distressed deteriorated	0.37348	0.4402	142	-1.23
Distressed no change	-0.04298	-0.02063	612	-1.19
Distressed recovered	-0.5831	-0.7751	190	4.19*
Not distressed	0.03121	0.09269	798	-2.48*
Not distressed improved	-0.2646	-0.4318	122	2.76*
Not distressed deteriorated	0.7459	0.4733	177	-4.65*
Not distressed no change	-0.03566	0.01192	498	-1.73
Beginning treatment	-0.12559	-.016657	685	1.53

Note. 5 cases could not be classified as distressed or not distressed because they were missing their initial OQ-45 score and were therefore excluded from subsequent analyses.

5 additional consumers could not have change statuses calculated because of missing data and were excluded from subsequent analyses.

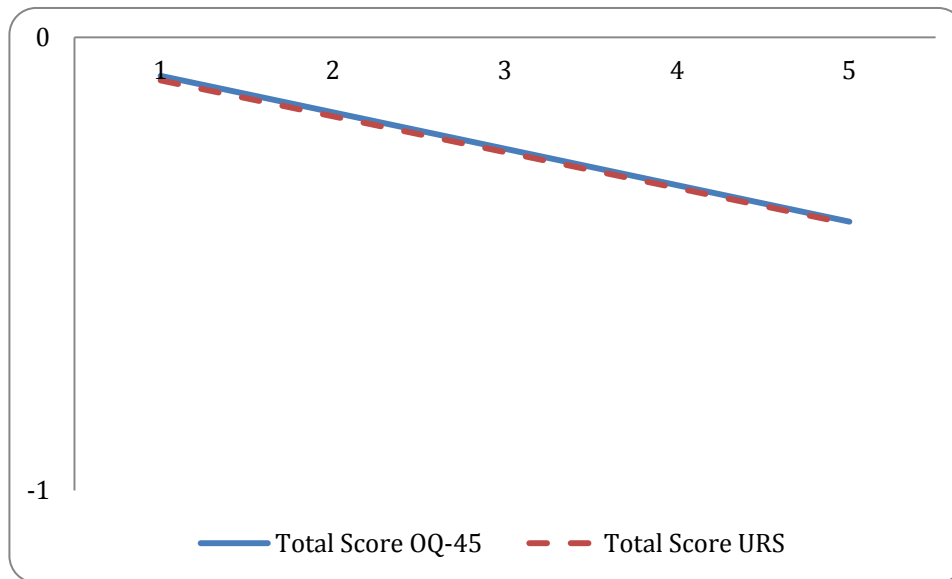
* $p < .05$.

Figure 1: Consumer URS total scores over time compared to zero.



	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.07654	0.01304	1093	5.87	0.0001

Figure 2: Consumer OQ-45 and URS total scores over time.



	Estimate	Standard Error	DF	T value	pr>t
Interaction	0.001246	0.01384	9080	0.09	0.9282