A Dissertation

entitled

A Rasch Analysis of the Mental Health and Recovery Measure: Reliability and Validity

by

Gunes Tanya Ozbey

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the

Doctoral Degree in Psychology

Wesley A. Bullock, PhD, Committee Chair

Svetlana Beltyukova, PhD, Committee Member

Jeanne H. Brockmyer, PhD, Committee Member

Stephen Christman, PhD, Committee Member

Janet M. Hoy, PhD, Committee Member

Patricia R. Komuniecki, PhD, Dean
College of Graduate Studies

The University of Toledo

December 2012
The treatment of serious mental illness has come a long way in just the last few decades, changing even what “mental health recovery” means. Parallel to this development, new measurements have evolved to capture the new meaning of recovery. The current study examined the reliability and validity, of a consumer-driven recovery measure, the Mental Health Recovery Measure (MHRM). The MHRM is a 30-item scale developed based on a grounded theory analysis (Young & Ensing, 1999) of information from individuals with severe and persistent mental illness regarding their understanding of the “lived experience” of recovery. The MHRM has previously been found (Bullock, 2005; Young, Ensing, & Bullock, 2000) to have good internal and test-retest reliability, convergent validity with other measures of recovery, and utility in the assessment of recovery following participation in recovery-focused treatment programs (Bullock, Sage, Hupp, O’Rourke, & Smith, 2009). The psychometric properties of the MHRM have been studied using classical statistical methods, which have limitations. The current study used the Rasch framework (Linacre, 2012a), which has significant advantages over the classical methods, including the ability to make sample-free examinations.
The current study used data from 1116 individuals with severe and persistent mental illness who attended the introductory session of the Wellness Management and Recovery (WMR) program in a variety of community mental health centers, consumer operated sites, vocational rehabilitation services and state psychiatric hospitals across Ohio. Analyses revealed that the MHRM has acceptable item and person reliability. However, examination of the response category diagnostics revealed that the 5-point item response scale was not used as expected by participant respondents. A modification of the response categories in which a 3-point response scale was developed revealed improved category functioning. The current 30 items comprising the MHRM appear to measure a single construct and the Rasch analyses did not support deletion of any items or shortening of the measure. The order of item difficulties was consistent with the recovery model; however, problems were identified regarding targeting people with very high and low levels of recovery and potential solutions to these problems are discussed.
To Nicolae, Ramona, Oana, Melanie, Amaro, and Teri: for the marks you left in my life. I sometimes wonder whether I came all the way across the ocean so that I could get to know you.
Acknowledgements

I am grateful to have Dr. Bullock as my mentor and clinical supervisor throughout my doctoral studies. His dedication and excellence as a teacher, a supervisor, and a therapist inspire me and remind me why I chose this profession. Furthermore, this dissertation would not have emerged without the help, support, inspiration, and encouragement from Dr. Beltyukova. I caught “the statistics virus” during one of her classes. Now I know I will always be a carrier of curiosity and the desire to extract the truth out of the data. I cannot thank you enough for your generosity and patience.

I would like to thank all of the members of my dissertation committee: Dr. Wesley Bullock, Dr. Svetlana Beltyukova, Dr. Sallyann Treadaway, Dr. Jeanne Brockmyer, Dr. Janet Hoy, and Dr. Stephen Christman for their genuine interest in the topic, for their support and guidance throughout the study. I also would like to thank Drs. Greg Meyer and Joni Mihura for their contribution to my skills and knowledge as a researcher and clinician.

I also would like to extent my gratitude to two extraordinary women in Turkey: Aydan and Fusun. I feel their unconditional love every day despite the distance.
Table of Contents

Abstract ........................................................................................................................................ iii

Acknowledgements ...................................................................................................................... vi

Table of Contents ....................................................................................................................... vii

List of Tables ............................................................................................................................... ix

List of Figures ............................................................................................................................. x

I. Chapter One: Introduction ...................................................................................................... 1

II. Chapter Two: Review of the Literature ................................................................................. 3

   A. History of the Treatment of Severe Mental Illness and Recovery .............................. 3
   B. Evidence Based Practices ......................................................................................... 9
   C. Recovery Models ....................................................................................................... 11
   D. Measuring Recovery ................................................................................................. 14
   E. Mental Health Recovery Measure ........................................................................ 19
   F. Rasch Analysis .......................................................................................................... 23
   G. Statement of the Problem ......................................................................................... 28
   H. Purpose of the Present Study ................................................................................... 29
   I. Research Questions .................................................................................................... 30

III. Chapter Three: Method ........................................................................................................ 31

   A. Setting ......................................................................................................................... 31
List of Tables

Table 1. Original Response Scale Diagnostic Statistics ........................................ 44
Table 2 Results of the Rasch Principal Components Analysis of Residuals. ........ 46
Table 3 Comparison of the Original 30-Item MHRM and Scale Alternatives ....... 53
Table 4 Scale Y Response Category Diagnostic Statistics ................................. 54
Table 5. Rasch Principal Components Analysis Results ...................................... 60
List of Figures

Figure 1. Original Response Scale Category Probability Curve....................... 45
Figure 2. Original 30-Item MHRM Scale Item Map........................................ 49
Figure 3. Scale Y Response Scale Category Probability Curve ....................... 55
Chapter One

Introduction

The discussion of recovery, including what it is, what it includes, and what term best describes it, continues among consumers, providers, and researchers across the world (Ralph, Kidder & Phillips, 2000). The definition of the construct has changed considerably over the last few decades in particular – changes which parallel the dramatic changes in the field of healthcare (Schultz, Chao & McGinnis, 2009). As the construct of recovery transforms, so does the measurement of it. One of the major transformations in the measurement of recovery has been the increasing shift to a consumer-point-of-view as opposed to an external/clinical/provider’s assessment.

The current study evaluated the reliability and validity of a consumer-driven measure: the Mental Health Recovery Measure (MHRM). Archival data was obtained from over 1000 individuals with serious mental illness who received psychiatric services in the State of Ohio. Reliable and valid measurement of recovery from mental illness is an important part of both treatment planning and treatment implementation. Furthermore, evaluating and improving the efficacy of the treatment programs is only possible by measuring the treatment outcome. The MHRM is being used across the United States as well as internationally (Andresen, Caputi & Oades, 2010). To date, the psychometric properties of the MHRM have been studied using classical statistical methods, which have limitations. For example, the results obtained by these methods are not sample-free. The current study used the Rasch framework (Linacre, 2012a), which has significant advantages over the classical methods, including the ability to make sample-free examinations. Reliability and validity diagnostics of the scale were studied using Rasch
analysis of the available archival data on the MRHM. The current study was designed to meet the need to address measurement problems in this area of research, not only with regard to the specific psychometric performance of the MHRM, but also with regard to how the results of the Rasch analysis of the MRHM speak more generally to the challenges of how to best measure the construct of recovery in a reliable and valid way.

The following review provides an overview of the literature on mental health recovery. It covers the history and current conceptualization of this construct, as well as definitional controversies. Additionally, measurement challenges are discussed. Lastly, an introduction to the Rasch measurement model is presented to provide a context for the subsequent methodology, statistical analyses, and results.
Chapter Two

Review of the Literature

History of the Treatment of Severe Mental Illness and Recovery

The treatment of serious mental illness has come a long way in the last four centuries. Across the years, the treatment model that was applied largely depended on the understanding of the nature of mental illness. For example, in the seventeenth century the emphasis was more about “what the illness is” rather than the treatment of the illness. Persons with severe mental illness were perceived as a potential threat to the society, and it was recommended that they be treated within the family (Grob, 1973). Institutional care began to be used starting in the eighteenth century and in the United States, the Pennsylvania Hospital was opened in 1752 in Philadelphia. The intended goal was to provide people with mental illness a place to heal and prepare to return to the community. However, the reality did not match this goal. The patients lived in damp, unheated cells with poor hygiene. In these early institutions, even what was delivered as “treatment” was harsh and punitive and the use of physical restraint was common. Other states proposed to establish institutions during the late 1700s, however few actually came into existence until the hospital movement in the 1800s (1973).

The mental health recovery concept as it is known today can be considered to be rooted in the Moral Treatment approach introduced by Philippe Pinel in France during the late eighteenth century. Pinel believed that mental illness was curable. The Moral Treatment approach was humanistic and emphasized that mental illness is a treatable condition. This notion was roughly based in a philosophy of recovery (Grob, 1973). The Moral treatment approach contributed to the increase in the number of private and public
asylums. Although Pinel believed that institutions were a necessary environment for the treatment of mental illness, he thought that people should be treated with patience, guidance, and dignity (Reisman, 1991).

The principles of Moral Treatment were spread to America by the early nineteenth century. These principles, along with social and economic changes of the time, contributed to the rise of public hospitals that could accommodate the individuals in need. However it was through the efforts of dedicated reformers that the principles of Moral Treatment were put in practice in these hospitals. One of these individuals was Dorthea Dix. She worked towards proper care and adequate facilities for people with mental illness, reducing the number of mentally ill individuals in jails, and expanding overcrowded hospitals (Grob, 1973). Dix’s efforts were continued by Clifford Beers in the 1900s. Beers initiated the “mental hygiene” movement, which aimed to improve the conditions in psychiatric hospitals, as well as the mental health of the community as a whole (U.S. Department of Mental Health, 1999). More state hospitals were established during the late 19th and early 20th century. To conclude, during the period from 1920’s to the middle of 1950’s, the standard for care for persons with severe mental illness was institutionalization. Although there were continuous efforts to improve the conditions in psychiatric hospitals, they were overcrowded, under or poorly staffed, and quality of care was routinely inadequate (Wegner, 1990).

The National Institute of Mental Health (NIMH) was first authorized by the U.S. Government in 1946, when then President Harry Truman signed the National Mental Health Act. NIMH, formally established in 1949, emphasized deinstitutionalization and the need for alternatives to treat mental illness. The focus of treatment shifted from
inpatient facilities to community-based settings. Both the number of people who stayed in the hospitals and the average duration of stay decreased. The understanding of mental illness shifted to a curable condition (U.S. Department of Mental Health, 1999).

Unfortunately, the community mental health system was not prepared to deliver adequate treatment to people who were discharged from the hospitals. Moreover, the interventions focused primarily on symptom reduction. The needs of these patients to live fruitful lives, with equal attention given to social, vocational, residential, and educational aspects of their lives, was not addressed (Anthony, 1993).

Even as efforts were being made during the 1950’s to improve mental health treatment at state institutions and move towards community-based care, powerful psychotropic medications were introduced as the new standard of care for mental illnesses (Wegner, 1990) and the medical model of mental health continued to predominate. The medical model perceived mental illnesses as comparable to chronic physical diseases. A traditional medical model of treatment continued to be used in the area of mental illness and medical professionals were in control. The causes of serious mental health problems were considered to be biological forces that were outside the person’s control, with an anticipated long-term course of unremitting deterioration. The medical model approach to treatment of mental illness emphasized limitations, weaknesses, deterioration, and external locus of control. Recovery was defined in negative terms (Davidson, Flanagan, Roe, & Styron, 2006; Schmook, 1996; Walsh, 1986). However, personal stories challenged the belief that mental illness is destiny (Deegan, 1988; Leete, 1989; Lovejoy, 1982).
In addition to these personal accounts, a social movement began in the 1960s among former mental health patients. The medical model’s assumptions that mental illness is chronic and the patients are dependent on medicine for the rest of their lives were challenged. This movement was called the “consumer-survivor” movement. These self-identified “mental health consumers” and “psychiatric survivors” emphasized empowerment of the individual instead of forced and involuntary treatment. It was suggested that consumer-survivors have an active role in designing and providing treatment services. The term consumer started to replace patient. The powerful role of language labels and the question of how best to refer to individuals receiving mental health services (e.g., “patients,” “clients,” “survivors,” “consumers,” “peers,” or “people receiving services”) are still unsettled. (Frese & Davis, 1997; McLean, 2003).

The medical model defined recovery as an endpoint to be reached (Resnick, Fontana, Lehman & Rosenheck, 2005). It implied that the goal of mental health recovery is a significant reduction or complete remission of symptoms, typically through the use of psychotropic medications or medical interventions (e.g., electroconvulsive therapy). However, several authors have pointed out that many individuals with severe mental illness did not deteriorate as predicted (Drake & Cotton, 1986; Drake et al., 2006; Harding et al., 1987; Harding, Zubin & Strauss, 1992), or that symptom remission per se was an inadequate definition of recovery. Moreover, the consumer perspective on recovery is different than that of mental health professionals (Davidson et al., 2005; Deegan, 1988).

Mental health consumers have defined recovery as a subjective, on-going journey that is more than an end point marked with decreased symptoms (Allott, Loganathan &
Fulford, 2002; Deegan, 1988; Frese & Davis, 1997; Ralph, 2000; Schmook, 1996). The definition of recovery in the President's New Freedom Commission report, Achieving the Promise: Transforming Mental Health Care in America (2003, p.7) emphasized that recovery is a process “in which people are able to live, work, learn, and participate fully in their communities.”

While there is no single definition of recovery, certain themes of recovery emerge across the literature. These recovery themes include self-efficacy, empowerment, and self-esteem (Bullock, Ensing, Alloy, & Weddle, 2000); hope and motivation (Allott, et al. 2002; Davidson et al. 2001; Deegan, 1996; Fisher, 1994; Young & Ensing, 1999); control over life (Anthony, 1993; Lovejoy, 1982, Young & Ensing, 1999); living a meaningful life (Anthony, 1993, Davidson et al., 2001; Young & Ensing, 1999); and symptom management (Deegan, 1994; Gingerich & Mueser, 2005). Although definitions of recovery usually identify recovery as a process rather than as a final end point, measures of recovery still typically focus on outcomes (Resnick et al., 2005). Thus, although recovery is thought of as an on-going process, it can also be treated as an outcome.

The consumer movement also led to another change in the mental health field. Consumers started to create and staff their own consumer operated centers, which are referred to as Consumer Operated Services (COS) centers. These COS centers can be as simple as a drop-in center that is run on a “club house” model with few formal activities or services, but may also include formal services such as case management services, self-help services, or specific peer support groups that are offered through the COS center (Bond, 1994; Chamberlin et al., 1996; Mead, Hilton & Curtis, 2001; Swarbrick, 2007).
Another contemporary concept in mental health services has been multicultural competency (US Department of Health and Human Services, SAMHSA, 2001). The Department of Health and Human Services released the report “Mental Health: Culture, Race, and Ethnicity” as a supplement to “A Report of the Surgeon General” (U.S. Department of Health and Human Services, 1999). The supplement gave directives for action in the areas of expanding the science base, improving access to mental health treatment and reducing barriers, improving quality of mental health services, supporting capacity development, and promoting mental health. To improve the quality of care for minorities, providers are encouraged to deliver culturally competent assessment and intervention services based on evidence-based professional guidelines. A culturally competent recovery program emphasizes the unique culture and background of each individual and provides a safe environment that respects cultural values, is free of racism, homophobia, sexism, and cultural imperialism. Cultural imperialism is the imposition of one culture over another and can take various forms, such as use of media images, societal pressure or a deliberate policy.

To summarize, the last century witnessed dramatic changes in the field of healthcare, which changed the understanding of mental illness and recovery (National Research Council, 2009). When proposing a definition of recovery in December, 2011, SAMSHA emphasized that it was still a “working definition.” This definition of recovery included four major dimensions of recovery: 1) health, 2) home (a stable and safe place to live), 3) purpose in life, and 4) community (relationships and social networks).
Evidence Based Practices

Another important movement in the mental health field has been the emergence of evidence based treatments. To be considered an evidence based practice (EBP), a practice must have been proven efficacious (and preferably effective) in producing expected outcomes in clinical trials, research settings, and real world settings such as community mental health centers (Hyde, Falls, Morris & Schoenwald, 2003). A practice is considered efficacious if there is evidence from at least two designs (randomized or quasi-experimental design studies, single-subject design studies) that the practice is superior to no treatment. This approach emphasizes that research should guide clinical services (Anthony, Rogers & Farkas, 2003; Frese, Stanley, Kress & Vogel-Scibilia, 2001). While “efficacy” refers to the outcomes under ideal or close to ideal conditions (i.e. laboratory), “effectiveness” refers to producing expected outcomes in “real life” settings. For example, an efficacy study might exclude people with comorbid disorders from a study, while in the real world of public health services as actually provided the consumers of the treatment might often have comorbid disorders.

Four key principles should be followed in evidence based mental health services: 1) Scientific evidence should be used in decision making, and this evidence should be provided to the consumers, 2) revising scientific evidence for the particular consumer and the circumstances, 3) consumers should be given the right to make decisions about their treatment based on accurate information about their illness, available treatments, expected outcomes and side effects, as well as their personal values, 4) practitioners should have the necessary expertise to provide the EBP (Drake, Rosenberg, Teague, Bartels, & Torrey, 2003).
There has been a vigorous debate about the compatibility of EBPs and the recovery model. Evidence based practices rely on empirical evidence regarding the effectiveness of treatment. The framework for EBP in mental health is based on the medical model, which focuses on symptom reduction (Anthony, 2001). Thus, evidence of effectiveness is often defined and measured in terms of symptom reduction or reduced use of treatment services.

In the view of some observers, EBPs and recovery-based approaches to mental illness constitute incompatible world views. Moreover, some observers contend that EBPs may serve to maintain traditional ways of providing services and thus pose the threat of reversing the advances made in promoting recovery (Anthony, 2001). Others have suggested that recovery approaches and EBPs can assist each other (Frese, Stanley, Kress & Vogel-Scibilia, 2001). For example, both models emphasize person-centeredness. Additionally, recovery can be used to identify the practices that will support the clinical process (Hyde et al, 2003). In their recovery model, Jacobson and Greeley (2001) suggested a way to link internal components of recovery (e.g. hope, healing, and empowerment) with external conditions (e.g. human rights, positive culture of healing, and recovery oriented services). Frese et al. (2001) proposed a model for integrating EBPs and the recovery model. The authors suggested that, when integrating the two, it is important to consider the degree of disability of consumers.

In the past decade, evidence based recovery programs have become as focus of policy makers, researchers and providers (The Substance Abuse and Mental Health Services Administration, 2012). The need to make EBPs more available for the treatment of severe mental illness was highlighted in the mental health report of the Surgeon
General (U.S. Department of Health and Human Services, 2001). The Substance Abuse and Mental Health Services Administration (SAMHSA) introduced the New Hampshire-Dartmouth National Implementing Evidence Based Practices Project to specifically address to that need. The purpose of this project was to promote the mental health interventions that have strong empirical support (Mueser et al., 2003). The following six practices were included in this project: Pharmacological Treatment, Assertive Community Treatment, Family Psychoeducation, Supported Employment, Integrated Dual Disorders Treatment, and Illness Management and Recovery (Mueser et al., 2006). Implementation resource kits were developed for each of these EBPs to foster their dissemination. The Illness Management and Recovery (IMR) program was developed based on research on teaching illness self-management strategies to individuals with severe mental illness. Five empirically supported strategies were incorporated into the program: psychoeducation about mental illness and its treatment, cognitive-behavioral strategies for medication adherence, developing plan for relapse prevention, social skills training, and coping skills training. In order to motivate clients to learn how to better manage their illness and to help them move forward in their lives, IMR begins with an exploration of the meaning of recovery to the client and setting personal recovery goals to work toward in the program (Mueser et al., 2003).

**Recovery Models**

In an effort to understand, predict and provide evidence for recovery, several models have been developed. These models provide a construct or a conceptual framework through descriptions, images, and definitions. Contemporary recovery models
have characteristics in common, but also bring different ideas to the concept and
definition of recovery (Andresen, Caputi & Oades, 2010; Ralph, et al., 2000).

DeMasi et al. (1996) proposed a well-being recovery model based on a review of
the extant literature. The three domains were identified: (1) physical and mental health,(2) psychological components, such as self-esteem, hope, coping, and confidence, and (3)social/quality of life components. During the same period, The Empowerment Model of
Recovery (Fisher & Ahern, 1999) was developed by the National Empowerment Center.
This model emphasized the importance of social supports in the recovery process.

Dornan, Felton and Carpinello (2000) developed another model: A Public Health
Model for the Recovery of Adult Mental Health. Hope was the central theme of this
model. Phases or characteristics of recovery focused on the major concept of hope,
including related concepts such as the will to survive, awakening, action plan, self and
shared determination.

Another recovery model is “The Medicine Wheel.” This is a relational worldview
model with origins in the American Indian tribal cultures. The Medicine Wheel appears
as a circle or sphere with four quadrants called context, mind, body, and spirit. The
quadrants represent the four major forces that together must come into balance (Cross,
Earle, Solie & Manness, 2000).

Other authors have proposed different models of recovery. For example, Jacobson
and Greenly (2001) proposed that the recovery process includes internal components
such as hope, healing, and connecting to others; and external components, such as human
rights, a supportive environment and recovery focused services.
In addition to the models that describe components of recovery or intrapersonal and external processes, there is a substantial literature that describes recovery as taking place in stages or phases. For example, Davidson and Strauss (1992) identified four phases of recovery:

(1) discovering the possibility of possessing a more active sense of self, (2) taking stock of strengths and weaknesses and assessing possibilities for change, (3) putting into action some aspects of the self and integrating the results as reflecting one's actual capabilities and (4) using an enhanced sense of self to provide some refuge to provide a resource against the effects of the illness and [such things as stigma] (p.134).

Baxter and Diehl (1998) proposed a three-stage recovery model: (1) Recuperation, (a time of dependence following crisis), (2) Rebuilding (a time of building independence), and (3) Awakening (a time of building interdependence).

The Recovery Advisory Group Recovery Model (Ralph and The Recovery Advisory Group, 1999) is another stage model created by a group of consumer leaders. The model defined Recovery through a number of stages: anguish, awareness, insight, action plan, determination to be well, and well-being/recovery. The model included both internal (within oneself) and external (interactions with others) factors. The authors of the model argued that the stages of the model do not form a linear path, and that people may move back and forth among the various stages.

The idea that the path of recovery is not linear has also been suggested by others (Andresen, et al., 2010; Ralph et al., 2000). Davidson and Strauss (1992) pointed out that
the stages of recovery do not necessarily occur in a linear fashion, but that there is a logical order to the four aspects.

Prochaska, DiClemente and Norcross (1992) developed a model of recovery from addiction that included a spiral four-step process of pre-contemplation, contemplation, preparation and action. They suggested that in the pre-contemplation stage, there is no recognition of the problem or intention to change. The contemplation stage is characterized by thinking about doing something about the problem as well as the intention, preparation and action to deal with the problem. The fourth step (action) is reached when the individual modifies his/her behavior in order to solve the problem.

Young and Ensing (1999) proposed a three-phase model of recovery. The first phase consists of acceptance, motivation to change, hope and inspiration. The second phase consists of self-empowerment, learning and self-redefinition, and achieving basic functioning. The last stage includes a sense of well-being and striving for higher potentials. Spaniol, Wewiorski, Gagne and Anthony (2002) said that recovery can be conceptualized in the following phases: being overwhelmed by the disability, struggling, living with the disability, and living beyond the disability.

**Measuring Recovery**

The task of evaluating recovery from mental illness has not been easy for healthcare professionals. Clinical judgment alone may not be enough. Because clinicians are not embedded into the natural environment of the consumers, they may not fully understand how the consumer functions in the day-to-day world. Therefore, clinicians rely on measures for assessing recovery. These measures may be used prior to treatment to help develop a treatment plan, during treatment to determine progress, and at the
conclusion of treatment to evaluate readiness for discharge or follow-up plan. Given that there are multiple definitions of recovery, it is not surprising that there are different measures that are based on different definitions and models of recovery. In this section, some of the contemporary recovery assessment measures currently being used will be reviewed.

The Test of Adaptive Behavior in Schizophrenia (TABS) (Velligan et al., 2007) was developed in 2007. For the TABS, a healthcare professional rates the adaptive behavior of the consumer to assess functional capacity in two domains: 1) initiation of tasks in the real world and 2) identifying problems that are faced during the course of functional activities. The TABS has 5 subscales: 1) Medication Management Skills 2) Empty Bathroom (i.e., identifying the items are needed in an empty bathroom for self-care), 3) Shopping Skills 4) Clothes Closet (i.e., how to select appropriate clothes for different circumstances), 5) Work and Productivity, and 6) Social Skills. Participants receive a “percent correct” score for each of the 6 domains. Total score is the average of the 6 domain scores and ranges from 0 to 100. The TABS demonstrated sufficient 3-month test-retest reliability with a Pearson coefficient of .8. Furthermore, internal consistency was good, with a Cronbach alpha for all items reported as .84. The TABS score was also related to other measures of functional outcome, with significant correlations being found between the TABS and the Multnomah Community Ability Scale (.49), and the Independent Living Skills Scales (.58). When scores on the TABS were compared between persons with schizophrenia and the controls group, those with schizophrenia demonstrated significantly worse performance (2007).
Another contemporary measure of recovery is the University of California, San Diego Performance-Based Skills Assessment (UPSA) (Harvey & Bellack, 2009). The UPSA was developed in 2001 as a measure of basic everyday living skills in elderly with schizophrenia. This measure is completed by a rater who evaluates the skills of the consumer. It consists of 5 subtests evaluating abilities in the following domains: 1) Planning/ Organization (e.g., planning a trip to the mall), 2) Finances (e.g. writing a check), 3) Communication (e.g. making medical appointments), 4) Travel (e.g. using a bus route schedule), and 5) Household (e.g. cooking). Total score ranges from 0 to 20 for each of the 5 domains. Total score ranges from 0 to 100. UPSA has shown to have sufficient interrater reliability ($r=.91$) and two-week test-retest reliability ($r=.93$) (Harvey, Velligan & Bellack, 2007). Furthermore, convergent validity (Petterson et al., 2006) and criterion validity (Twamley et al., 2002) have been demonstrated for UPSA.

Another measure that assesses recovery is The Connor-Davidson Resilience Scale (CD-RISC) (Connor & Davidson, 2003). The CD-RISC is a 25-item self-report scale. The content of the scale includes control, commitment, change, developing strategies with a clear goal, action orientation, coping, humor in the face of stress, stable bonds, social problem solving skills, patience, stress endurance, and faith. Each item is on a 5-point scale where 0 indicates “not true at all” and 4 indicates “true nearly all of the time.” The total score can range from 0 to 100, with higher scores indicating greater resilience. CD-RISC measure has been found to have good internal-consistency, test-retest reliability, and content validity (Connor & Davidson, 2003).

Another measure of recovery is the Illness Management and Recovery-Client Self Rating Scale (IMR Scale), which was developed to evaluate the outcome of participation
in the IMR program. The IMR Scale contains 15 items rated on a 5-point behaviorally anchored scale. The IMR Scale was developed to assess the different components of illness self-management taught in the IMR program (Gingerich & Mueser, 2005). Specifically, the IMR Scale includes items that assess progress toward goals, knowledge about mental illness, involvement with significant others and self-help, time in structured roles, impairment in functioning, symptom distress and coping, relapse prevention and hospitalizations (Mueser et al, 2006). Both the clinician and consumer versions of the IMR scale have demonstrated adequate internal reliability, with the consumer version yielding coefficient alphas of .68 and .72 at time 1 and time 2, respectively. The clinician version yielded alphas of .80 at time 1 and .71 at time 2. Pearson correlations between the responses at time 1 and time 2 indicated good 2-week test-retest (.81 for both versions). The IMR Scale also demonstrated sufficient convergent validity (Salyers, Godfrey, Mueser, & Shauna Labriola, 2007).

Some of currently used measures of recovery were reviewed in this section. As seen above, these measures of recovery are based on each author’s theory of what “recovery” is or what it should look like (e.g. reduction in symptoms or increase in adaptive functioning). Even the method of administration of the measure implies the approach of the theory behind it. For example, the UPSA and the TABS are administered by a professional, whereas the IMR Client Self Rating Scale and The Connor-Davidson Resilience Scale are completed by the consumer, which is congruent with an approach that empowers the consumers. Furthermore, the items in the UPSA and the TABS are based on external indicators as opposed to their internal experience (e.g. whether or not the consumer uses their medication as prescribed, as opposed to whether or they are
comfortable with their use of medication). The IMR Client Self Rating Scale, on the other hand, is an example of an outcome measure that is based on the content of a specific psychoeducational curriculum, that is, it actually assesses how well the consumer learned and applied the content of the program. A common limitation of many current measures of recovery is that they were developed through a top-down approach, using clinical “expert” knowledge, the treatment goals of the particular clinical intervention, or the researcher’s own theory as starting and ending point. In that case, it is uncertain whether any of these measures cover the domains of recovery that have been identified as important by the consumers themselves as “experts” on the recovery process, since they have the “lived experience” of recovery to draw from (Deegan, 1988).

Recovery measures are used to measure the impact of specific interventions, to measure the progression of recovery over time, and to evaluate whether changes are related to the usage of mental health services (Ralph et al., 2000). However, problems were noted regarding contemporary measures’ ability to measure change (Andresen et al., 2010; Ralph et al., 2000). In their review of recovery measures, Ralph et al. (2000) noted that limited information was available regarding the utility of recovery measures to measure change over time or related to an intervention. The review also pointed out that one of the reviewed measures, The Crisis Hostel Healing Scale (New York Crisis Hostel Project, 1998) did show significant change over time in a study that used an experimental group. Ralph et al. (2000) also pointed out that when a recovery measure does not cover the whole construct of recovery, it is not possible to discuss an ability to measure change. Furthermore, the authors pointed out that the items in contemporary recovery measures concentrated on the positive side of recovery. These authors suggested that in order for
the measurements to truly mirror the range of recovery, they should include anguish and similar aspects of recovery (Ralph et al., 2000).

Andresen et al. (2010) reviewed contemporary recovery measures in relation to the clinical measures. The authors concluded that the change of subcomponents of recovery might be dependent upon the stage of recovery. For example, for the Mental Health Recovery Measure with 41 items (Young & Bullock, 2003) on six subdomains (overcoming stuckness (6 items); discovering and fostering self-empowerment (6 items), learning and self-redefinition (9 items), return to basic functioning (6 items); striving to attain overall well-being (6 items) striving to reach new potentials (8 items)), highly significant stage effects were found across five stages of recovery as defined by the authors. More specifically, overcoming stuckness was relatively level across the first four stages and decreased at stage 5. This subdomain included a variety of themes, including willingness to work hard on recovery, acceptance of illness, spirituality and seeking help from others. Furthermore, the scores for the empowerment domain increased across stages, but was quite level between stage 2 (Awareness) and stage 3 (Preparation). Scores for all other MHRM subdomains increased across five stages. These results supported a stage model of recovery and indicated that the measure of recovery might be dependent upon the associated stage.

**Mental Health Recovery Measure**

The Mental Health Recovery Measure (MHRM) (see Appendix A) was developed by Young and Bullock (2003), as a 30 item self-report measure of mental health recovery. The items comprising the MHRM were developed using a bottom-up approach, that is, from a grounded theory model of recovery that was based on the lived experience
of recovery by individuals with mental illness (Young & Ensing, 1999). Young and Ensing (1999) interviewed 18 individuals with severe and persistent mental health problems, with diagnoses ranging from Bipolar Disorder to Borderline Personality Disorder. Grounded theory analysis was used to identify common, underlying components of the mental health recovery process. The outline of the resulting recovery model is provided in Appendix B. As a consequence, the MHRM is specifically designed to evaluate recovery domains that have been identified as important by the consumers themselves, such as getting “unstuck,” spirituality, redefining oneself, learning, or developing new potentials. Therefore, the MHRM has a holistic and consumer-driven approach to assessing the mental health recovery process. Additionally, the MHRM is a self-report scale, thus, the consumers respond to the items according to their own perceptions, as opposed to using external evaluators or external indicators.

The initial MHRM was developed based on the themes from the Young and Ensing (1999) study and had 36 items. Initial psychometric analysis of the measure was completed by Young (1999), when the MHRM was used as an outcome measure in a treatment program for consumers with psychiatric illness. In this study, the overall scale had a high level of internal consistency (alpha=.91). Correlations between the MHRM and other measures that have been used to assess the recovery process were used to evaluate the convergent validity of the MHRM, particularly, the Empowerment (“Making Decisions”) Questionnaire (Rogers, Chamberlin, Ellison, & Crean, 1997) and the Community Living Skills Scale (Smith & Ford, 1990). The results demonstrated that the MHRM Total Score had a significant positive correlation with both the Community Living Skills Scale ($r=.746$) and the Empowerment Questionnaire ($r=.523$) (Young,
Reliability of the MHRM was demonstrated by a sample of 279 consumers from five community mental health centers and two consumer-operated service centers. The average total MHRM score in this sample was 80 (SD=20). The internal reliability (coefficient alpha) of the scale was .93. Additionally, one week test-retest reliability for this sample was .92. In a subsample of 180 consumers, the convergent validity was calculated with two subscales of the Empowerment Questionnaire, with correlations of .67 and .73, respectively (Bullock, 2005).

Later, five more items were added to the 36-item MHRM scale to better assess the recovery domains of New Potentials, and Learning and Self-Redefinition (Bullock, 2005). Finally, the 41 item version of the MHRM was analyzed using internal reliability analyses comparing item to total reliability and was item total was reduced to the current 30 items (Bullock).

Subsequent studies continued to demonstrate the internal reliability and convergent validity of the MHRM. In a sample of 150 mental health consumers from two community mental health centers, significant positive correlations with established measures of recovery were demonstrated,. The correlation was .67 with the Empowerment Scale (Rogers, Chamberlin, Ellison, & Crean, 1997), .73 with the Conner-Davidson Resilience Scale (Connor & Davidson, 2003), and .75 with the Resiliency Scale (Wagnild & Young, 1993) (Breedlove, 2005). Additionally, the correlation with the Community Living Scale (Smith & Ford, 1990) was .57 in a sample of 180 (Bullock & Young, 2003).

In the most recent sample (Bullock, 2009), between 433 and 514 consumers completed the MHRM and other recovery measures. The MHRM Total Score was
significantly negatively correlated ($r=-.45$), with the Symptom Distress scale on the Ohio Consumer Outcomes Measure (Ohio Department of Mental Health, 2000), that is, as the recovery score increased, symptom distress decreased. Additionally, the MHRM Total Score was significantly correlated with the Safety and Health Scale ($r=.39$). No significant relationship was found with the Quality of Life Scale ($r=.09$). The MHRM Total Score was also significantly positively correlated with the Empowerment Scale ($r=.58$) (Rogers et al., 1997). Among the subscales of the Empowerment Scale, the Total MHRM Score was correlated most highly with the Self Esteem ($r=.68$) and Optimism ($r=.45$) subscales. The Total MHRM Score was modestly correlated with the Community Activism subscale ($r=.29$). The MHRM Total Score was not significantly correlated with the Power-Powerlessness ($r=.05$) or the Righteous Anger ($r=-.17$) subscales.

Regarding reliability, with a recent sample of 671 people from 17 community mental health centers, the coefficient alpha of the MHRM items was .95. The average Total MHRM score for this sample was 78 (SD=21.7) (Bullock, 2009).

As mentioned below, the content of MHRM is based on the information gathered from mental health consumers (Young & Ensing, 1999). The items, “I eat nutritious meals everyday,” and “I go out and participate in enjoyable activities every week,” exemplifies nature of the MHRM items. The measure has eight conceptual domains: 1) Overcoming Stuckness (items 1, 2, 3, and 4), 2) Self-Empowerment (items 5, 6, 7, and 8), 3) Learning and Self-Redefinition (items 9,10,11, and 12), 4) Basic Functioning (items 13, 14, 15, and 16), 5) Overall Well-Being (items 17, 18, 19, and 20), 6) Spirituality (items 25 and 26), 7) New Potentials (items 21,22,23, and 24), and 8) Advocacy/Quality of Life (items 27,28,29, and 30).
Individuals respond to each MHRM item on a 5-point Likert scale, ranging from “Strongly Disagree” which corresponds to a 0, to “Strongly Agree” which corresponds to a 4. The MHRM Total score is derived by adding up the numbers corresponding to the response of each item and ranges from 0 to 120. Higher scores correspond to a higher self-reported level of recovery.

Rasch Analysis

A key concern in psychometrics is whether scores are consistent, that is reproducible across different test situations. This has led to the development of several indices of reliability depending on test administration design: internal consistency, rater consistency, test-retest reliability, and alternate-forms reliability (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999).

Reliability can also be defined as the degree to which test scores are free from measurement error and accounts for error due to sampling. The value of reliability is theoretically bounded by zero and one and reflects the inter-item correlation among item responses. A single test administration yields an estimate of reliability as below:

\[
\text{Reliability} = \frac{\text{true variance}}{\text{observed variance}}
\]

and, when measures and errors are uncorrelated,

\[
\text{Observed variance} = \text{true variance} + \text{error variance}
\]

To illustrate the problems with measurement of reliability using classical methodological techniques, one of the measures for reliability is Kuder-Richardson 20, which is defined as below:
\[ \alpha = \frac{K}{K - 1} \left[ 1 - \frac{\sum_{i=1}^{K} p_i q_i}{\sigma_X^2} \right] \]

where \( K \) is the number of items and where the variance for \( \sigma_X^2 \) in the denominator is:

\[
\sigma_X^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}
\]

And \( p_i \) is the proportion of correct answers on item \( i \), and \( q_i \) is the proportion of the incorrect answers. Each \( p_i q_i \) is the variance of responses to an item \( i \) for a person from whom \( p \)-value is the probability of succeeding (i.e. obtaining a score of “1”) on item \( i \). In other words, each \( p \) value is what is expected from the average person from that sample response variance for a particular item. The summation of \( p_i q_i \) yields the score variance for a person who has those \( p \) values. However, persons with low or high scores have less error variance than persons who have scores near 50%. In summary, a significant methodological weakness of using the average person in KR20 will be the overestimation the error variance of high and low scorers. Furthermore, in typical research studies, estimates of internal consistency are based on previous applications of the assessment. However, a formal estimate is not informative unless the samples have exactly the same score distributions. Finally, using raw scores to calculate the sample variance is potentially misleading as raw scores are non-linear (Schumacker, 2007; Smith, 2001).

The problems with measurement of reliability with the classical theory can be addressed with Rasch analysis. Rasch model is based on fit theory. According to this theory, there are two sources of data present: one represents the real world (i.e. observed phenomenon) that the researcher wishes to investigate; the other is the mathematical model that is thought to represent the real world. When the data is analyzed using a
Rasch model, an estimate is obtained about what our construct might be like if we were to create a ruler to measure it. The Rasch model provides a useful approximation of the measure that helps to understand the process underlying the reason why people and items behave in a particular way. After obtaining representative data about the observed phenomenon, Rash analysis can be carried out to find out the goodness of fit. If the data fit the Rasch model, they can then be interpreted in terms of equal-interval units using log transformations of raw data odds and probabilistic equations. In this way, one can determine the extent to which the items have consistently measured a single variable from easy to difficult in a monotonically increasing fashion (Bond & Fox, 2007).

The Rasch model utilizes two variables: item difficulty and person ability. The difficulty and ability values are real numbers without measurement unit. Item difficulty refers to the endorsability of the item. Person ability refers to the extent the person embodies the measured construct. Their meaning becomes evident when the conditional probability of each element of the item response matrix is introduced. Some researchers also refer to the scale of ability and difficulty as a logarithmic measure called “logit,” which expresses that ability and difficulty are on the same scale. Rasch analysis assists in the evaluation of the constructed metric, and the extent to which the items in the measure function unidimensionally. Items in each scale should measure a common underlying construct otherwise it is not appropriate to combine them to generate a scale score. Useful measurement also requires an equal-interval scale and a hierarchical arrangement of the items along the variable continuum where some items are meaningfully less difficult than others. Using Rasch fit statistics it is possible to examine the contribution of each item to the measurement of a single construct by assessing the extent to which an item or an
individual performs as expected. If the model has a good fit, easy items are endorsed by more respondents than are difficult items. Additionally, respondents who have more of the measured construct would endorse more of the “difficult” items than respondents who have less of the measured construct. A good scale would successfully measure the intended construct with only necessary and relevant items (Smith, 2001).

Rasch analysis can be used to assess reliability. The “item reliability index” refers to the replicability of item placements along the pathway if these same items were given to another sample with comparable ability levels. In other words, if other persons were given these same items, would the item estimates remain stable? If high item reliability is achieved, then the scale has some items that are more difficult and some items that are easier, and these inferences are consistent.

Rasch models can address the problems of classical theory that are noted above, because if the data fit the model, each person’s ability and each item’s difficulty are on a linear scale and thus suitable for the calculations of means and variances (Schumacker, 2007; Smith, 2001). Therefore, Rasch models provide a direct estimation of the modeled error variance for each estimate of a person’s ability and an item’s difficulty. These standard errors provide quantification of the precision of each person’s measure and item difficulty. In this way, the range within each item’s true difficulty or a person’s true ability falls can be described. Individual standard errors are more useful than a sample or test average, which overestimates the error score variance of persons with high and low scores. A group estimate of reliability can be produced by summing the squares of individuals’ standard errors, as opposed to using the error variance of the average person sampled. This produces a better estimate of internal consistency because the numerical
values are now linear if the data fit the model, and actual average error of variance of the sample is used instead of the error variance of an average person (Schumacker, 2007).

Another crucial question about the scales has been how to determine the appropriate number of rating scale categories. Research has suggested that the addition of response categories might increase reliability only if these additional categories are not arbitrary (Linacre, 1995; Wright & Linacre, 1992). The increase in the number of response alternatives might introduce error by allowing respondents to draw more freely on divergent frames of reference. In such a situation, it is difficult to establish a common language between the respondent and the investigator via the rating scale. In short, two people might perceive the same level of anxiety, yet one may check a 5 and the other a 6, simply because the introduction of too many response options muddles the definition of the variable in question.

The challenge is that there is no definitive optimal number of response categories that applies to all rating scales. Five response categories might work for accurately measuring one construct, yet a simple yes-or-no type of response might be best for another. Therefore, every time a new rating scale is developed, the optimal number of response categories must be determined empirically. Determining the optimal number of response categories requires examination of Rasch measurement diagnostics. This is achieved by analyzing the estimated difficulty of choosing one response over another (e.g. “strongly agree” over “agree,”). Thresholds should increase monotonically and should be appropriately distanced from each other (i.e., should be at least 1.4 logits, but not farther than 5 logits apart) (Linacre, 1999). If any problems are revealed in the existing rating scale, the number of response options can be reduced by collapsing
problematic categories with adjacent, better-functioning categories. Next the data can be reanalyzed. Diagnostics from the new analysis can be compared with those from the original analysis to see whether the collapsing helped to improve variable definition. The aim is to produce the rating scale that yields the highest quality measures for the construct of interest (Bond & Fox, 2007).

To summarize, Rasch analysis can be used to assess internal consistency and reliability and has advantages over the classical measures of reliability. Furthermore, Rasch analysis provides an evaluation of the contribution of each item to the scale. In this way, an optimum scale can be established where each item contributes to the measurement by differentiating people with regard to the extent that they possess the measured construct. Rasch analysis can also be used to determine the optimum number of response categories. The validity and reliability of the MHRM scale were demonstrated using traditional methods. Rasch analysis provided a better estimate of reliability and gave more insight into construct validity and scale category functioning.

**Statement of the Problem**

The concept of recovery from serious mental illness has received increasing amounts of attention in the fields of mental health and mental health care (New Freedom Commission on Mental Health, 2003). The development of the concept of recovery has been concurrent with the recognition that our methods and models of clinical service delivery should be guided by empirically-based outcomes research. Given the stated goal of the national public mental health agenda and the need to create and empirically evaluate clinical interventions that are specifically designed to promote the recovery process, it is critical to have a reliable and valid measure of mental health recovery.
Having a reliable and valid recovery measure is not only important for program evaluation and outcome research purposes, but is an important component of treatment planning, implementation and follow-up.

The MHRM is a 30-item self-report questionnaire of recovery that is based on a grounded theory analysis of interviews with individuals with severe and persistent mental illness. The MHRM has demonstrated satisfactory reliability and validity with traditional methods. However, traditional methods of reliability have limitations. Rasch analysis can provide a more accurate estimate of reliability. Moreover, Rasch analysis can be used to obtain construct validity by exploring unidimensionality and ranking of items by difficulty. Finally, Rasch analysis can indicate whether the MHRM has the optimum number of response categories, and can also be used to reduce the number of categories if the current number is not optimum.

**Purpose of the Present Study**

The purpose of the present was to analyze archival MHRM data, using Rasch analysis, to gather improved evidence for reliability and validity of the measure and to indicate and address those elements of the scale that negatively impact validity and reliability of the MHRM. The following aspects of validity were examined: 1) whether the measure’s response categories were used appropriately, 2) whether the items of the scale were on one single dimension, and 3) whether the items spread along the scale from easy items to difficult ones.
Research Questions

Research questions were used in lieu of specific research hypotheses, as they were more appropriate for the purposes of the present study:

(1) Does the MHRM Scale have a satisfactory reliability?

(2) Does the MHRM have the optimum number of response categories?

(3) Do the MHRM items measure a single underlying variable?

(4) Does the item difficulty hierarchy make sense in the context of the recovery model on which it is based?
Chapter Three

Method

Setting

Current study used archival MHRM data that was collected from individuals who attended the introductory session of the 10-week Wellness Management and Recovery (WMR) group in the community. WMR was developed in part from Illness Management and Recovery (IMR). The theory behind WMR is based on recovery and wellness. The program uses a holistic approach to promote happier and healthier lifestyles. It is structured and curriculum based. WMR is implemented at sites across State of Ohio in rural and large urban community mental health centers, consumer operated sites, vocational rehabilitation services and state psychiatric hospitals. The Wellness Management and Recovery Coordinating Center of Excellence (WMR CCOE) coordinates WMR and is supported by the Ohio Department of Mental Health (ODMH). Current sites include community mental health centers such as Southeast, Inc. in Franklin County, Greater Cincinnati Behavioral Health Services in Hamilton County, Neighboring Inc. in Lake County, and Daymont Behavioral Health Services in Montgomery County, as well as consumer operated sites such as Bridges Mental Health Consumer Empowerment, Gathering Hope House in Lorain County, and The Recovery Center of Hamilton County (Wellness Management and Recovery Website, 2012). The WMR program is delivered while psychiatric services, substance use and case management are being provided by a community agency. The program is delivered in a group format in
ten two-hour sessions (The Wellness Management and Recovery Coordinating Center of Excellence (WMR CCOE), 2006).

One of the essential characteristics of WMR is the consumer’s active role in his or her own recovery. In WMR, consumers take ownership of their wellness and are perceived as the experts of their own experience of the mental illness. The consumers’ personal choices are respected. With the guidance of the facilitators, the participants set personal goals for themselves and work towards them. The consumers are included as peers in the program and collaborators in the planning for services (WMR CCOE, 2006). Additionally, five of the current WMR implementation sites are consumer-operated sites and run entirely by consumers (WMR Website, 2010).

Another central value of WMR is cultural competency. WMR groups respect the unique culture of their participants. Individual beliefs and values are accepted as a resource. WMR also acknowledges the subjective nature of the recovery process. In addition to each person having a unique background, WMR also acknowledges that each participant will have unique goals and face different challenges to reach their personal goals.

WMR utilizes four mental health intervention approaches that were adapted from IMR: (1) Psychoeducation, (2) Cognitive-Behavioral Methods for using Medication Effectively, (3) Relapse Prevention, and (4) Coping Skills Training. Research supports that these components facilitate recovery (Mueser et al., 2002). The curriculum consists of ten areas: 1) Meaning of recovery, 2) Understanding mental health, 3) The role of medication in recovery and wellness, 4) Learning to manage symptoms and side effects, 5) Effective communication, 6) Communicating with your providers, 7) Wellness, 8)
Coordinating your care, 9) Building social supports and involving others, and 10) Planning for wellness (WMR CCOE, 2006).

Monitoring the outcome of the program is an integral part of the WMR program. Quantitative and qualitative recovery outcome data from individual participants are collected before (pre-data) and after (post-data) they complete the program, and six months after the completion of the program (follow up-data). Two self-report measures are used: 1) The Mental Health Recovery Measure (MHRM; Young & Bullock, 2003), 2) the WMR Client Self-Rating Scale (Mueser, et al., 2004). The Mental Health Recovery Measure (MHRM; Young & Bullock, 2003; Bullock, 2005) was not specifically developed for WMR, but was designed to measure the recovery process for individuals with serious and persistent mental illnesses. Currently, the MHRM is being used as a clinical program outcome measure at 19 WMR implementation sites across the State of Ohio.

**Participants**

The current study used archival MHRM data obtained from 1116 individuals with serious mental illness who received psychiatric services in the community in the State of Ohio. The sample constituted all the pre-treatment data obtained from participating sites by March 2012. Individuals in this sample had diagnoses that included affective disorders (e.g., bipolar disorder) and psychotic-spectrum disorders (e.g., schizophrenia). The average age of the participants was 45.4 years (range 20.4 to 71.7). Gender and ethnicity information was obtained from Ohio Mental Health Consumer Outcomes System-Adult Consumer Form (Ohio Department of Mental Health, 2000) that used to be part of the
data collection in the WMR program. However, during the data collection phase, this form was discontinued by Ohio Department of Mental Health. As a result, in the final sample, a large percentage of gender and ethnicity information was missing. The percentage of people who reported their gender was 25.5%. Among the participants who reported their gender, 54.5% are female. Regarding ethnicity, 18 % of the participants identified their ethnicity as one of the four categories: 80 % of the participants were European Americans, 10 % of the participants were African American, 5.2% were Latino, and 3.3% identified as Native American/Pacific Islander. None of the participants identified themselves as Asian American, although that was one of the options.

Measures

**Mental Health Recovery Measure.** The Mental Health Recovery Measure (MHRM) is a 30 item self-report measure of mental health recovery. Individuals respond to each MHRM item on a 5-point Likert scale, ranging from “Strongly Disagree” which corresponds to a 0, to “Strongly Agree” which corresponds to a 4. The MHRM Total score is calculated by adding up the numbers corresponding to the response of each item and ranges from 0 to 120. Higher scores correspond to a higher self-reported level of recovery. The MHRM is administered before the participant attends the WMR program (pre-MHRM), right after the program (post-MHRM), and 3-6 months following the completion (follow-up MHRM).

Data Analysis

The demographics were obtained using SPSS (Statistical Package for the Social Sciences), version 20.0. WINSTEPS (Linacre, 2012b) computer software, version 3.75
was used for the Rasch analyses. Analyses were determined in accordance with research questions; specific analyses and acceptable values or cut-offs are discussed below. It is important to emphasize that when using Rasch analysis to construct a measure, decisions are rarely made based on only statistical evidence (Bond & Fox, 2007). Therefore, the acceptable and cut-off values are better conceptualized as guidelines that were used in conjunction with guiding theory.

**Research Question One.** In order to address the first research question, person and item reliability indices as well as person and item separation indices were consulted using WINSTEPS (Linacre, 2012b). The person reliability index and the item reliability index are indicators of the overall stability of the resulting person and item hierarchies. The person reliability index can be defined as the extent to which the sample respondents would fall in the same order if given a different measure of recovery. The item reliability index indicates the extent to which the hierarchy of items (from “easiest”/most frequently endorsed to “hardest”/least frequently endorsed) would remain stable if the same items were administered to another sample of equal size. When these indices are below acceptable values, common explanations are that either the measure is lacking breadth in item content or that the sample of respondents did not include individuals falling within the full range of recovery from mental illness. The person reliability index is comparable to Classical Test Theory reliability coefficient Cronbach’s alpha. Item reliability is unique to Rasch analysis. These Rasch reliability indices are on a scale from 0 to 1 (Bond & Fox, 2007). The closer the reliability index is to 1, the greater the stability of the person and item hierarchies. For the current study, acceptable value for person reliability indices was set at 0.80 and above (Bond & Fox, 2007). Acceptable value for item
reliability indices was set at 0.90 and above (Linacre, 2012a). For the present study, these values were re-evaluated after modifications were made to the measure in order to ensure that attempts to increase validity did not negatively impact reliability to a significant degree.

The Rasch item separation statistic for the entire instrument was used to determine the level of distinction possible among items and persons. Separation is the ratio of the square root of the variance explained by the measurement model to that of the unexplained variance or measurement error, including error from model misfit. If person separation is smaller than 2 in a relevant sample and person reliability is smaller than .80, the measure may not be sensitive enough to distinguish between high and low performers (Bond & Fox, 2007). More items may be needed (Linacre, 2012a). If the item separation is smaller than 3 and item reliability is smaller than .90, the implication is that the items are not distinguishable in terms of “low, medium, high” item difficulties. In this case, the person sample may not be large enough to confirm the construct validity of the instrument (Linacre, 2012a).

**Research Question Two.** The functioning of the rating scale categories was examined using Rasch rating scale diagnostics. The following data was analyzed: rating scale category frequencies, average rating scale category measures, unweighted mean square fit statistics, rating scale category thresholds, and a visual assessment of probability curves.

The rating scale category frequencies, in other words observed count, provide the raw frequencies of category utilization in the overall sample. To qualify as a meaningful response category, each category must have been utilized frequently enough (i.e. at least
10 observations) (Linacre, 2012a). Low frequencies on any of the categories indicate that the category structure might be problematic and some categories might need to be collapsed with another category (Bond & Fox, 2007). The average measure is the average “ability” for all participants in the sample that endorsed that rating. These values should increase as rating scale categories increase. Categories that deviate from this expected pattern are also considered candidates for collapsing with adjacent categories.

Fit statistics and structure calibrations (i.e., “thresholds”) are empirical indicators of category functioning. Outfit mean-square statistics are required to be less than 2.0 in order to be consistent with Rasch model expectations (Linacre, 1999). Thresholds indicate the difficulty estimated in choosing one response category over the previous category (e.g., difficulty in endorsing “Strongly Agree” rather than “Agree”). It is anticipated that these values increase as rating scale categories increase if each category provides a unique and meaningful difference from other categories. In addition to directionality, the magnitude of the distance between these threshold values is also important. The thresholds should be neither too close nor too far apart. The present study used recommended guidelines that thresholds should increase by between 1.4 logits and 5 logits (Linacre, 1999). When the distance between neighboring categories is less than 1.4, the interpretation is that the distinction between rating scale categories is not sufficient. Thus these categories are candidates for collapsing. If the distance between any two categories is greater than 5 logits, adding an additional category is considered.

Lastly, category probability curves were consulted to further examine category functioning (Linacre, 2012b). The probability curves indicate the probability of endorsing each rating scale category, given “X” difference between item “difficulty” and person
“ability”. If all rating scale categories function properly, the figure reveals each category demonstrates a distinct peak “as a distinct hill” (Linacre, 1999, p. 115). If categories overlap and collapse in the middle, the interpretation is that these categories are not useful in determining meaningful differences in levels of the latent construct. Once the categories that are not used in a consistently distinct and informative manner are identified; they can be collapsed (i.e. re-coded) to improve reliability and interpretation of the measure.

**Research Question Three.** Item fit, person fit and analysis of residuals were used to address research question three. Item-measure (point-biserial) correlations were consulted as negative or very low correlations might indicate items that contradict the overall content of the measure (Linacre, 1995). Item and person fit statistics (infit and outfit) were used to identify any items or participants that misfit, in other words confound the interpretation of the variable. Items or persons with Infit or Outfit statistics that were outside of the acceptable range were candidates for removal from the analysis. Additionally, the possibility of any underlying patterns to the misfitting items and participants were investigated.

The guideline by Linacre (2012a) was used to assess item fit. Fit statistics greater than 2.0 indicates off-variable noise that is greater than useful information. Such an item would degrade the instrument. Fit statistics between 1.5 and 2.0 indicates noise that neither constructs nor degrades measurement. Linacre (2012a) interprets fit statistics between .5 and 1.5 as an indicator that the item is productive for measurement. Linacre (2012a) suggested that infit or outfit mean squares below < .5 indicates overly predictable item. Items that underfit the model might be measuring a construct other than the
construct of interest. Similarly, person misfit mean square values above 2.0 are interpreted to indicate those people who endorsed items unpredictably, considering the item hierarchy (e.g., endorsing harder/less frequently endorsed items, while failing to endorse easier/more frequently endorsed items) (Bond & Fox, 2007). The Rasch model contains some expectations of variation. Therefore, it is acceptable for up to 10% of the sample to demonstrate significant “misfit.” These instances could be attributed to carelessness or inattention, item bias, or response sets. However, if a greater proportion of the sample demonstrates clear misfit, other issues, such as multidimensionality, may be the reason.

If data fit the model perfectly, the primary factor would explain all (i.e., 100%) observed variance and the residuals would not be correlated. However, perfect adherence is not expected. Some amount of unexplained variance is typically observed. As long as the majority of the variance is explained by the measure and that additional factors identified within the residuals are minimal, it can be assumed that the measure has a unidimensional construct (Bond & Fox, 2007).

Rasch principal components analysis (PCA) was used in this analysis to assess unidimensionality. In Rasch PCA, the primary factor (i.e., the factor that is explained by the measure) is extracted from the data and then the principal component analysis is conducted on the remaining standardized residuals. This analysis detects significant patterns (i.e., correlations) within the variance that is not explained by the measure. If the total percent of variance explained by the measure is below 50%, further analysis is necessary in order to assess for the presence of meaningful patterns within the residual variance (Linacre, 2012a). If the amount of unexplained variance is greater than 5% in
the largest contrast and any of the secondary components have an Eigenvalue greater than 2.0, then the measure may not be unidimensional (Bond & Fox, 2007; Linacre, 2012a). When the possibility of multidimensionality is indicated, factor loadings needs to be examined to determine the items that are loading on the significant non-primary components. Items with strong positive (.40 and above) or negative (-.40 or below) loadings on these factors can be studied for theoretical relations in order to determine if the former items are better suited for a separate scale. Standardized residual correlations will also help to investigate item redundancy and to determine if it is possible to shorten the MHRM without significant loss of information. The proposed plan for the study was that if any items were removed, the ability of the new item set to quantify the difference of post-MHRM and pre-MHRM scores would be compared to the 30-item version, using the post-treatment data. At the end, this step was not completed as all 30-items were retained based on the results of the Rasch analysis and the results indicated problems with item hierarchy, which would need to be resolved to ensure accurate measurement of post-pre differences.

**Research Question Four.** Additional information about validity was obtained by analyzing the item difficulty hierarchy. Useful measurement requires an equal-interval scale and a hierarchical arrangement of the items along the variable continuum where some items are meaningfully less difficult than others. The Rasch model transforms the raw data into equal interval units called logits by applying log transformations and probabilistic equations, within standard error estimates for each person and each item (Bond & Fox, 2007). The Rasch person-item map displays a ruler created from the measurements of the individuals’ responses to the MHRM items, in other words their
level of progress on each recovery item, and the progress needed to get high scores on each item. When the person-item map indicates that the items do not spell out a meaningful construct from easiest to hardest, common explanations are that either the measure is lacking breadth in item content (i.e., the full range of recovery) or that the sample of respondents does not include individuals falling within the full range of recovery.

**Limitations**

One limitation of the study was related to sampling and consequent limitations regarding the generalizability of the results. The data was collected from 1116 participants who presented for the introduction session of a 10-week psychoeducation program for people with severe and persistent mental illness in the community. A total of 19 agencies administered the measure. Agencies might have had different group referral procedures which might have affected the sampling of participants. Moreover, participation in these groups is voluntary. Therefore, the people who have decided to attend to this program might be somewhat different than people who had a severe and persistent mental illness, but chose not to start, or chose not to complete the program and hence were lost to post-test and other follow-up. However, this limitation is less problematic in Rasch framework as compared to traditional methods due to separate estimations for item and person parameters.

Another limitation of the study was the self-report format of the measure. Several disadvantages were noted for the self-report method. One was the honesty and accuracy of answers by each respondent. The participants’ responses had to be accepted at face value with regards to thoughtfulness and honesty. In most self-report measures, several
motives might distort the accuracy of individuals’ responses. These motives include consistency seeking, self-enhancement, and self-presentation (Paulhus & Vazire, 2009). Choosing Rasch methodology addressed the limitations of self-report method to some extent as Rasch analysis enables the researcher to identify inconsistencies in responses.
Chapter Four

Results

This chapter is divided into two sections. In the first section, results from the Rasch analysis of the 30-item MHRM scale using the Pre-treatment data are presented and problems are identified. The second section includes the steps for addressing these problems and the associated results.

Section One: Rasch Analysis of the Original 30 Item MHRM

Reliability. The Rasch analysis revealed a reliability of .91 (SE=.03) for persons and .99 (SE=.07) for items. Person separation was 3.19 and item separation was 10.88. All the values were greater than the minimum acceptable values referenced in Chapter Three. Thus, the 30-item MHRM instrument revealed adequate reliability.

Category Functioning. The diagnostic analysis of the original 5-point response scale summarized in Table 1 (0 = “Strongly Disagree”, 1 = “Disagree”, 2 = “Not sure”, 3 = “Agree”, 4 = “Strongly Agree”) indicated that the infit and outfit mean squares were within acceptable range for all categories, that is, did not exceed 1.5 (Linacre, 2012). Furthermore, each category had at least 10 occurrences, and sample and expected averages increased as rating scale categories increased. However, the distances between the first four categories were less than 1.4 logits. In other words, these categories were close to each other. And thus these categories were identified as candidates for combining.
Table 1.

*Original Response Scale Diagnostic Statistics*

<table>
<thead>
<tr>
<th>Category Label</th>
<th>Category %</th>
<th>Sample Expected</th>
<th>Observed Av.</th>
<th>Infit MNSQ</th>
<th>Outfit MNSQ</th>
<th>Thresholds</th>
<th>Category Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>0, “Strongly Disagree”</td>
<td>8</td>
<td>-.66</td>
<td>-.57</td>
<td>1.15</td>
<td>1.60</td>
<td>None</td>
<td>-2.41</td>
</tr>
<tr>
<td>1, “Disagree”</td>
<td>13</td>
<td>-2.23</td>
<td>-2.22</td>
<td>.99</td>
<td>1.11</td>
<td>-1.01</td>
<td>-1.04</td>
</tr>
<tr>
<td>2, “Not sure”</td>
<td>18</td>
<td>.19</td>
<td>.15</td>
<td>.89</td>
<td>.97</td>
<td>-.34</td>
<td>-.15</td>
</tr>
<tr>
<td>3, “Agree”</td>
<td>38</td>
<td>.70</td>
<td>.66</td>
<td>.93</td>
<td>.86</td>
<td>-.30</td>
<td>.94</td>
</tr>
<tr>
<td>4, “Strongly Agree”</td>
<td>22</td>
<td>1.69</td>
<td>1.76</td>
<td>.99</td>
<td>.99</td>
<td>1.65</td>
<td>2.85</td>
</tr>
</tbody>
</table>

*Note:* MNSQ = Mean Square. Av = Average.

Figure 1 is a visual representation of category functioning. The probability curves illustrate the probability of endorsing each rating scale category, given a difference between person “ability” and item “difficulty.” The figure also revealed that the categories 1=Disagree and 2=Not Sure did not have a distinct peak, in other words they did not emerge “as a distinct hill” (Linacre, 1999, p. 115). This indicates that at no point one of these categories was the most probable response for a part of the measure.
Figure 1. Original Response Scale Category Probability Curve

**Dimensionality.** The Rasch principal components analysis of residuals revealed that the variance explained by the 30-item measure was 43.2%, which was less than the acceptable minimum of 50% (Linacre, 2012a). As seen in Table 2, analysis of contrasts indicated that the largest secondary dimension explained more than 5% of the variance. The eigenvalue of 2.7 was somewhat bigger than 2, the smallest amount that could be considered a "dimension" (Bond & Fox, 2007). However, the percentage variance explained by item difficulties (24.1%) is more than 4 times of the variance explained by the second dimension (5.2%). Thus the primary dimension was dominant (Linacre, 2012a)
Table 2

Results of the Rasch Principal Components Analysis of Residuals

<table>
<thead>
<tr>
<th></th>
<th>Standardized Residual Variance (Eigenvalue Units)</th>
<th>Variance Observed (%)</th>
<th>Variance Modeled (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Raw Variance in Observations</td>
<td>52.8</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Raw Variance explained by measures</td>
<td>22.8</td>
<td>43.2%</td>
<td>43.5%</td>
</tr>
<tr>
<td>Raw Variance explained by persons</td>
<td>10.1</td>
<td>19.1%</td>
<td>19.2</td>
</tr>
<tr>
<td>Raw Variance explained by items</td>
<td>12.7</td>
<td>24.1%</td>
<td>24.3%</td>
</tr>
<tr>
<td>Unexplained variance (total)</td>
<td>30.0</td>
<td>56.8%</td>
<td>56.5%</td>
</tr>
<tr>
<td>Unexplained variance in the 1st contrast</td>
<td>2.7</td>
<td>5.2%</td>
<td>9.1%</td>
</tr>
<tr>
<td>Unexplained variance in the 2nd contrast</td>
<td>2.1</td>
<td>4.1%</td>
<td>7.1%</td>
</tr>
<tr>
<td>Unexplained variance in the 3rd contrast</td>
<td>1.9</td>
<td>3.6%</td>
<td>6.4%</td>
</tr>
<tr>
<td>Unexplained variance in the 4th contrast</td>
<td>1.7</td>
<td>3.1%</td>
<td>5.5%</td>
</tr>
<tr>
<td>Unexplained variance in the 5th contrast</td>
<td>1.5</td>
<td>2.9%</td>
<td>5.1%</td>
</tr>
</tbody>
</table>

**Item hierarchy.** The item map was examined to see if the hierarchy of the items was meaningful and consistent with the theory. As seen in the item map (Figure 2), item 30 ("I have enough money to spend on extra things or activities that enrich my life,")) was the “most difficult” item. In other words, it was likely to be endorsed by persons who had a higher level of mental health recovery and unlikely to be endorsed by persons who were
at the initial stage of recovery (Young & Ensing, 1999). The word “extra” implies that one has the money above and beyond what is needed to “survive.” As discussed in the literature review, several models of recovery suggested that mental health recovery occurs in stages. These stages are not mutually exclusive. Although the conceptual boundaries of the stages might differ between models, there is an agreement that increasing overall quality of life (e.g., having extra money) is characteristics of a more advanced stage of recovery. To sum, the location of item 30 at the top of the item hierarchy was consistent with the theory.

The easiest items to endorse were item 1 (“I work hard towards my mental health recovery,”) and item 9 (“Every day is a new opportunity for learning,”). Both items were associated with the initial stage of recovery from mental illness.

Additionally, the items in the very middle (items 4 and 17) were examined. Item 4 was, “I take risks to move forward with my recovery.” According to model proposed by Young and Ensing (1999), having the courage to challenge self and take risks falls into the middle phase of recovery (“Regaining and Moving Forward.”). Item 17 was, “I feel good about myself.” Being patient with self and discovering new aspects of self were also themes of the middle phase of recovery (Young & Ensing, 1999). Therefore, the positions of these items in the item hierarchy were consistent with the theory.

As seen in Figure 2, further indicated that there was a gap in the hierarchy between item 30 and the rest of the items. It is recommended that the items spread out more or less evenly in the item map for adequate targeting of persons with different abilities (i.e. different levels of recovery). Examining the end point of the item map revealed that the measure was not sensitive enough to sufficiently differentiate persons
with very high and very low abilities from others. The persons (as seen on the left side of
the item map) in the very high and very low end did not have any items that were aligned
in the associated levels. For example, there were more than 176 people who were
measured above 1 logit and more than 8 of them were measured above 6 logits. However,
the most difficult item (item 30) was measured at 1 logit. This indicated that more items
(difficult and easy) needed to be added to the measure.
Item polarity and misfit. Item polarity analysis did not reveal any items with negative or low correlation with the overall measure. Using the guidelines described in
Chapter Three, only Item 30, (“I have enough money to spend on extra things or activities that enrich my life,”) was identified as misfitting (infit mean square =1.70, outfit mean square = 2.02). As discussed above, item 30 is the most difficult item in the item hierarchy. However, it is not unusual for the most difficult and easiest items to misfit (Bond & Fox, 2007). Deletion of a misfitting item should be considered only if addressing other problems (e.g., poor category functioning) does not resolve the misfit and that the deletion of the item is consistent with the theory. Item 30 targets the financial aspect of standard of living and the ability to afford leisure activities (Young & Ensing, 1999). Deleting this item would not be consistent with the theory. Using the guidelines previously described, none of the items were identified as overfitting. Fit statistics for every item is provided in Appendix C.

**Person Misfit.** As the Rasch model contains some expectations of variation, it is acceptable for up to 10% of the sample to demonstrate significant “misfit.” Person fit statistics indicated that 11.1% of the participants had outfit mean square greater than 2. They were examined to see if they were different in terms of demographics, including the location of treatment (agency). No differences were found between this subgroup and the rest of the participants. Next, most unexpected responses were analyzed qualitatively. These unexpected and inconsistent responses were typically a score of “0” or “1” on an item while the expected response was more than “3,” based on item difficulty and the high scores that were endorsed on other items by the same person. Removing some of these participants from the data set might be necessary to decrease the noise in the results (Linacre, 2012a).
**Summary.** The analysis of the original 30-item scale revealed adequate person and item reliability and separation. Examining category functioning revealed that response categories did not function as expected. More specifically, the first four categories were not sufficiently distinct from each other. Additionally, the dimensionality analysis indicated that the measure explains less than 43.2% of the variance, which is less than adequate. Additionally, item 30 was misfitting. Although the item hierarchy was consistent with the stage model of recovery, the items were not evenly spread out. Most of the items were positioned around the average level with gaps towards both ends of the hierarchy, most significantly between the most difficult item and the rest of the items. Finally, 11.1% of the persons were identified as misfitting, which is higher than by chance. The steps taken to solve these problems are addressed in greater detail in the next section.

**Section Two: Optimizing Response Categories and Model Fit**

In this section, category functioning was improved by combining response categories “Strongly Disagree” and “Disagree” and dropping the “Not Sure” category. Thus a 3-point scale was obtained (0=Disagree, 1=Agree, 2= Strongly Agree). Next, the results was examined again using the new response categories for model fit by using mean square item and person fit statistics and Rasch principal components analysis. Using the same mean-square fit guidelines as before, and as a result of several iterations, 170 of the 1116 persons were deleted. Furthermore, 14 people were deleted because they endorsed the highest score on every item they responded to. At the end, the sample of 1116 was reduced to 932 people. The variance explained by the measures increased from
47.1% to 50.0%. The measure maintained acceptable person and item reliability and separation during these iterations.

**Category functioning:** To improve category functioning, two strategies were considered – category collapsing (where it was meaningful based on the conceptual similarity of the categories) and elimination of the category “Not Sure.”. The following combinations were empirically tested:

- **Scale X** - 4 response categories: “Strongly Disagree”, “Disagree”, “Agree” and “Strongly Agree”. “Not sure” response category was omitted (recoded as “missing”).
- **Scale Y** - 3 categories: “Disagree”, “Agree,” and “Strongly Agree.” Response categories “Strongly Disagree” and “Disagree,” were combined. “Not sure” response category was omitted (recoded as “missing”).
- **Scale Z** - 4 response categories: “Disagree,” “Not Sure,” “Agree,” and “Strongly Agree”. Response categories “Strongly Disagree” and “Disagree,” were combined.

Table 3 summarizes the results of the Rasch analysis of each rating scale modification. As seen in Table 3, Scale Y appeared to be the only scale that demonstrated acceptable category functioning. Furthermore, the variance explained by the measure increased from 43.2% to 47.1%. The category functioning diagnostics for Scale X and Z are provided in Appendix D and Appendix E. The category functioning diagnostics for Scale Y is provided in Table 4 and Figure 3.
Table 3 *Comparison of the Original 30-Item MHRM and Scale Alternatives X, Y, Z*

<table>
<thead>
<tr>
<th></th>
<th>Original 30-item MHRM</th>
<th>Scale X</th>
<th>Scale Y</th>
<th>Scale Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Persons</td>
<td>1116</td>
<td>1116</td>
<td>1116</td>
<td>1116</td>
</tr>
<tr>
<td>Number of Items</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Response Categories</td>
<td>5 (Strongly Disagree, Disagree, Not Sure, Agree, Strongly Agree)</td>
<td>4 (Strongly Disagree, Disagree, Agree, Strongly Agree)</td>
<td>3 (Disagree, Agree, Strongly Agree)</td>
<td>4 (Disagree, Not Sure, Agree, Strongly Agree)</td>
</tr>
<tr>
<td>Person Reliability</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td>Person Separation</td>
<td>3.19</td>
<td>3.12</td>
<td>3.19</td>
<td>3.29</td>
</tr>
<tr>
<td>Item Reliability</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Item Separation</td>
<td>10.88</td>
<td>9.63</td>
<td>9.75</td>
<td>10.85</td>
</tr>
<tr>
<td>Variance Explained by the Measure (%)</td>
<td>43.2</td>
<td>47.1</td>
<td>47.1</td>
<td>43.6</td>
</tr>
<tr>
<td>Category Functioning</td>
<td>-The threshold values increase by less 1.4 logits for first 4 cat.s -Cat. “1” and “2” do not reach a peak in the plot</td>
<td>-The threshold values increase by less than 1.4 logits for the first three cat. -Cat.”1” corresponds to “0” logit</td>
<td>Acceptable</td>
<td>-The threshold value decreases as response cat. increase from “1” to “2”. -Cat. “1” does not reach a peak and is not total distinct from the lowest and the highest cat.</td>
</tr>
<tr>
<td>Misfitting Items</td>
<td>30</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Overfitting Items</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Item Hierarchy</td>
<td>Items are not spread out. Range: -.85 to 1</td>
<td>Items are not spread out. Range: -1 to 1.5</td>
<td>Items are not spread out.</td>
<td>Items are not spread out. Range: -1.4 to 1.8</td>
</tr>
</tbody>
</table>
The response category functioning diagnostics of the 3-point scale (Scale Y) is presented in Table 4 and Figure 3 and revealed that the new response categories function as expected. Each category had a distinct peak that corresponded at least .5 probability or more. In other words, each response category was the most probable response at some point. The 3-point scale was used in the next step for the initial iterations to optimize model fit.

Table 4

*Scale Y Response Category Diagnostic Statistics*

<table>
<thead>
<tr>
<th>Category Label</th>
<th>Category %</th>
<th>Sample Expected</th>
<th>Observed Av.</th>
<th>Infit MNSQ</th>
<th>Outfit MNSQ</th>
<th>Thresholds</th>
<th>Category Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>0, “Disagree”</td>
<td>25</td>
<td>-1.52</td>
<td>-1.52</td>
<td>1.06</td>
<td>1.23</td>
<td>None</td>
<td>-2.51</td>
</tr>
<tr>
<td>1, “Agree”</td>
<td>46</td>
<td>.00</td>
<td>-.01</td>
<td>.85</td>
<td>.82</td>
<td>-1.36</td>
<td>.00</td>
</tr>
<tr>
<td>2, “Strongly Agree”</td>
<td>29</td>
<td>1.82</td>
<td>1.83</td>
<td>1.02</td>
<td>1.07</td>
<td>1.36</td>
<td>2.51</td>
</tr>
</tbody>
</table>

*Note: MNSQ= Mean Square. Av. =Average.*
Figure 3. Scale Y Response Scale Category Probability Curve

Optimizing Response Model Fit: First, the persons at the top in the person-item hierarchy (persons with the highest abilities) were examined. Examination of their responses revealed that 12 of these people endorsed the highest rating on all 30 items, thus obtained the maximum possible score (120). Additionally, two people endorsed the highest rating on some of the items and skipped (i.e., left blank) the rest of the items. These 14 persons were not identified as misfitting by the Rasch analysis because, mathematically, they responded as “predicted” based on their abilities. However, such scores were not consistent with the expectations for people who would take this measure before treatment. Although it is expected that people who are referred to the WMR program for the first time would be heterogeneous with regards to where they are in their recovery journey, the referral and acceptance to such a program also indicates that at least one area of improvement was identified for this person. This is consistent with the general philosophical definition of recovery as a never ending process rather than a
targeted, finite end point. Therefore, these 14 people were removed from the rest of the analysis.

The fit statistics calculated with 14 persons removed from the sample of 1116 persons revealed that the percentage of misfitting persons was 7.78% (87 persons). Using SPPS, no differences were found between this subgroup and the rest of the participants in terms of demographics and location. Next, these 87 people were analyzed separately with Rasch analysis. The reliability and separation indices revealed less than acceptable reliability. More specifically, person reliability and separation were .87 and 2.56 respectively. Item reliability and separation were .27 and .61, respectively. The unidimensionality analysis revealed that the percentage of variance explained by the measure was 38.8 %, which was below the minimum acceptable value. Person misfit analysis did not indicate any misfitting people. Item misfit analysis indicated that item 16 (“I am comfortable with my use of prescribed medications,”) was misfitting (infit mean square =1.46, outfit mean square = 2.28). The category functioning diagnostics revealed that the 3-point scale was not used appropriately, and indicated that the misfitting group may have used the response scale in a different manner as compared to the fitting people (comparison provided in Appendices F & G).

Lastly, item hierarchies for the misfitting versus fitting group were compared and were found to contain significant differences in item order. For example, although 30 was the most difficulty item in the original sample; in the misfitting sample, this item was one of the easier (than average) items in this sample (comparison provided in Appendices H & I).
Because item 16 was misfitting for this sample, it was deleted and the analysis was repeated to examine any improvement in model fit for the misfitting sample. Using 29 items with the same 87 people, person reliability and separation were .84 and 2.26 respectively. Item reliability and separation were .30 and .65, respectively. The unidimensionality analysis revealed that the percentage of variance explained by the measure is 38.7 %, which was below the minimum acceptable value. These results revealed that the 87 misfitting people were creating significant noise and were deleted from further analysis. Next, person misfit was analyzed again. It is not unusual for new persons to misfit after the initial misfitting people are deleted (Linacre, 2012a). Indeed, person misfit statistics indicated that 29 people were misfitting. These 29 people were then analyzed with the 87 people that were initially identified as misfitting. The reliability and separation indices revealed less than acceptable reliability. More specifically, person reliability and separation were .86 and 2.47 respectively. Item reliability and separation were .60 and 1.23, respectively. The unidimensionality analysis revealed that the percentage of variance explained by the measure is 37.0 %, which was below the minimum acceptable value. Lastly, item hierarchies for the misfitting versus fitting group were compared and were found to contain significant differences in item order.

It was decided that these 29 misfitting people created significant noise. Therefore they were removed from the analysis. Using the same mean-square fit guidelines as before and as a result of several iterations, 170 people were identified as misfitting, at which point the person fit statistics did not indicate more misfitting people. To understand any patterns in this sample, these 170 people were analyzed separately. The misfitting group demonstrated less than acceptable reliability (person reliability = .85,
item reliability = .85, person separation = 2.37, item separation = 2.39), compared to the fitting people statistics. Additionally, the response scale category functioning was impacted, indicating that the misfitting group may have used the response scale in a different manner. More specifically, response category “1” (Agree) did not reach a peak, thus this was never the most probable response at any point (comparison provided in Appendices J & L). Moreover, the category thresholds increased by less than 1.4 logits. The percentage of variance explained was 36.0%, which is below minimum acceptable level. Lastly, item hierarchies for the misfitting versus fitting group were compared and were found to contain significant differences in item order (comparison provided in Appendices K & M).

Lastly, item 16 (“I am comfortable with my use of prescribed medications,”) was misfitting. Deleting this item did not improve the results. More specifically, using 29 items (i.e. deleting item 16), reliability indices did not indicate acceptable reliability (person reliability = .84, item reliability = .86, person separation = 2.26, item separation = 2.44). The percentage of variance explained was 36.4%, which is below minimum acceptable level.

To conclude, it was decided to exclude these 170 persons from the rest of the analysis to decrease noise. When the data from the remaining 932 people was examined, the variance explained by the measures increased to 50.0 %, which is the minimum acceptable level for unidimensionality. Results of the principal components analysis are provided in Table 5. Analysis of contrasts indicated that the largest secondary dimension has an eigenvalue of 2.6, but this construct explained less than 5% of the variance.
Therefore, this construct did not constitute a significant second dimension. Analysis of standardized residual correlations did not indicate any dependent items.
Table 5

*Rasch Principal Components Analysis Results*

<table>
<thead>
<tr>
<th></th>
<th>Standardized Residual Variance (Eigenvalue Units)</th>
<th>Variance Observed (%)</th>
<th>Variance Modeled (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Raw Variance in Observations</td>
<td>60.0</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Raw Variance explained by measures</td>
<td>30.0</td>
<td>50.0%</td>
<td>49.8%</td>
</tr>
<tr>
<td>Raw Variance explained by persons</td>
<td>24.3</td>
<td>40.6%</td>
<td>40.5</td>
</tr>
<tr>
<td>Raw Variance explained by items</td>
<td>5.6</td>
<td>9.4%</td>
<td>9.4%</td>
</tr>
<tr>
<td>Unexplained variance (total)</td>
<td>30.0</td>
<td>50.0</td>
<td>50.2%</td>
</tr>
<tr>
<td>Unexplained variance in the 1st contrast</td>
<td>2.6</td>
<td>4.4%</td>
<td>8.8%</td>
</tr>
<tr>
<td>Unexplained variance in the 2nd contrast</td>
<td>1.9</td>
<td>3.2%</td>
<td>6.3%</td>
</tr>
<tr>
<td>Unexplained variance in the 3rd contrast</td>
<td>1.8</td>
<td>2.9%</td>
<td>5.9%</td>
</tr>
<tr>
<td>Unexplained variance in the 4th contrast</td>
<td>1.6</td>
<td>2.7%</td>
<td>5.4%</td>
</tr>
<tr>
<td>Unexplained variance in the 5th contrast</td>
<td>1.4</td>
<td>2.4%</td>
<td>4.8%</td>
</tr>
</tbody>
</table>

Next, item fit was analyzed again after the sample of people was reduced to 932 people. No misfitting items were identified. No overfitting items were identified. Final item fit statistics are provided in Appendix N.

**Reliability of the Transformed Scale:** The reliability and separation indices were examined again using the 3-point scale and 932 persons. The Rasch analysis
revealed a reliability of .92 ($SE=.06$) for persons, which is slightly higher than the initial value (.91). Item reliability remained the same at .99 ($SE=.15$) for items. Person separation and item separation were comparable to the values obtained in the initial analysis (i.e. 3.43 and 10.19, respectively).

**Item Hierarchy of the Transformed Scale.** Item hierarchy of the 3-point scale is provided in Appendix M. The hierarchy of the items was similar to the hierarchy revealed using the original 30-item scale and 1116 persons and was still consistent with the theory of recovery. Furthermore, the new distribution of item hierarchy had a larger range. More specifically, the range of item difficulties was -1.4 to 2 as compared to -.85 to 1 (the range in the initial analysis). There were fewer items that were on the same level of difficulty. However, there were still gaps between the most difficult item and the rest of the items as well as and the easiest item and the rest of the items. Examining the end point of the item map also revealed that the measure may not sufficiently differentiate person with very high and very low abilities from others. The persons in the very high and very low end do not have any items that are aligned in the associated levels.

**Summary.** Several steps were taken to address the problems identified by the Rasch analysis of the 30-item MHRM scale. However, not all of the identified problems were able to be resolved.

1. Problematic category functioning was resolved by combining the first two response categories and re-coding the “Not Sure” category as missing. The resulting 3-point scale (0=Disagree, 1=Agree, 2= Strongly Agree) functioned well.
2. Person misfit was addressed by removing some of the most extreme misfit cases. Additionally 14 people were deleted because they endorsed the highest rating on every item they have completed.

3. After the response scale was transformed and misfitting people were deleted, item misfit improved without deleting any items. The item fit statistics did not reveal any misfitting or overfitting items.

4. Item hierarchy somewhat improved after the response scale was transformed and most misfitting people were deleted and was still meaningful according to the theory of recovery. More specially, in the new map, there were more levels of item difficulty. However, there were still gaps at the two ends of the hierarchy. Additionally, the range of persons’ abilities was larger than the item difficulties on both ends. This indicates that the measure does not adequately differentiate people with very low and very high levels of recovery.

5. Unidimensionality improved after the response scale was transformed and most misfitting people were deleted. At the end, the variance explained by the measure was at the minimum acceptable level. No significant secondary dimension was identified. This was supportive of the unidimensionality of the measure.
Chapter Five

Summary and Discussion

Summary of the Study

Measuring recovery from mental illness is an important part of both treatment planning and treatment implementation. Furthermore, evaluating and improving the efficacy of the treatment programs is only possible by measuring the treatment outcome. Therefore, establishing relevant measures that are appropriately sensitive to changes in the recovery process for individuals is a crucial part of the treatment process.

The current study examined the reliability and validity of a measure of recovery, the MHRM. The results have conceptual and practical significance. Measuring recovery is essential to evaluate what treatments work under which conditions. Ignoring measurement issues would mean that the treatment decisions are made without having a good understanding of the construct of recovery. Objective and reproducible measures are unidimensional, have parameter separability, and an equal-interval hierarchical structure that is meaningful (consistent with the theoretical more/less continuum). Currently, the only framework that allows the examination of these aspects of a measure is the Rasch framework.

In the current study, first, the reliability of the scale was demonstrated by Rasch analysis, which has superiorities over classical test construction methods. Second, problems with item response category functioning were identified and addressed by combining the first two response categories and eliminating the middle category. The measure was also found to be unidimensional. Furthermore, problems with item hierarchy were identified. These results will enable the further improvement of the
measure, and our ability to understand the construct of recovery and how people make progress towards their recovery with treatment.

**Key Findings:**

This study evaluated the reliability and validity of a self-report measure of mental health recovery, the MHRM, using Rasch analysis. The reliability and separation indices revealed adequate reliability.

Despite the promising results with respect to reliability and separation, the MHRM’s 5-point response scale did not function in a meaningful way. Specifically, the distances between the first four categories were disordered. These results suggested that the participants did not use the scale in a monotonic fashion, which is necessary for an objective measurement. More specifically, the respondents were not able to differentiate between “Strongly Disagree” and “Disagree.” Moreover, the “Not Sure,” category was not distinct from neighbor categories on each end (“Disagree” and “Agree.”) Therefore, retaining all five categories was not justified. The “Not Sure” category did not present as a distinct category, even when the positive and negative categories are combined within themselves. Combining the first two response categories (“Strongly Disagree” and “Disagree,”) and re-coding the middle category “Not Sure” as a missing response improved category functioning. The result was an unbalanced scale (0= Strongly Disagree/Disagree, 1=Agree, 2= Strongly Agree). However, separating out agreement categories were supported by the results of the Rasch analysis. Both “Agree” and “Strongly Agree,” represent a different point on the continuum and the participants were not using them in an idiosyncratic manner. The impact of unbalanced scales on the quality of responses is an empirical question (Bond & Fox, 2007).
When person fit was examined, 11.1% of the sample demonstrated significant misfit, meaning that a significant portion of the sample endorsed items in a manner that was inconsistent with the identified hierarchical expectations (i.e., endorsed items higher on the hierarchy, without endorsing items lower on the hierarchy). Demographic variables were analyzed in order to compare the “misfitting” and “fitting” groups, and no noteworthy differences were observed with regard to gender and race/ethnicity representation or agency (where mental health services are provided) within each group. However, the interpretation of this finding is limited at this point due to missing demographical information. When misfitting people were analyzed separately using Rasch analysis, it was concluded that these people used the response scale in a different manner as compared to the fitting people. Moreover, the order of item difficulties was different in comparison to the fitting people. When reliability and separation indices were examined for misfitting people, less than acceptable reliability was revealed. Thus, the responses by the misfitting people were not reliably explained by a Rasch model. The inconsistency in responses might be related to a more general challenge of the measurement of recovery because the concept of recovery has not been clearly defined. The discussion of recovery, including what it is, what it includes, and what term best describes it, continues among consumers, providers, and researchers across the world (Ralph, Kidder & Phillips, 2000). It has even been suggested that recovery is such a personalized process that it is not possible or beneficial to produce a concrete definition (Ralph et al., 2002). The proponents of this view pointed out personal accounts as the best measure of recovery. The MHRM was developed on a model of recovery that was consumer-driven. Young and Ensing (1999) had developed this model through an
analysis of personal accounts of recovery. However, since recovery is a personal journey, not every consumer may find every aspect of the model relevant to their journey. Person misfit might also be related to confusion or inattentiveness. In the current study, the unexpected responses that were most likely to distort the measure were removed from further analysis. Furthermore, 14 people who endorsed the highest score on every item they completed were removed from the analysis. This decision was made based on the expectation that persons who were at pre-treatment stage for the WMR program have room for improvement in at least some of the domains of recovery. A total of 184 persons were removed. Item fit and unidimensionality (as indicated by Rasch principal components analysis) improved after the deletion of these people.

The present results provided evidence for several aspects of validity. First, after most misfitting persons were excluded, the principal components analysis indicated that the measure is unidimensional. Furthermore, the item fit statistics supported unidimensionality to a large extent. In the initial analysis, item 30 was misfitting. After the response categories were transformed and most unexpected responses were removed; no misfitting item was identified. To sum, the items were found to measure a single construct of recovery. Furthermore, the hierarchy of items indicated that the responses of participants demonstrated a logical hierarchical ordering. However, problems were identified regarding test targeting. The majority of items clustered in the middle. The appropriateness of item difficulties was compromised when person ability was very low or very high. People with very low and very high levels of recovery did not have any items that targeted at their level of recovery. This indicated that the measure was not
adequately differentiating those at the higher and lower levels of recovery from those who have abilities closer to the average range.

The MHRM was developed on a model of recovery that was developed by Young and Ensing (1999) based on information gathered from consumers. The MHRM items reflect the themes of recovery that were described in three phases of recovery. The model identified significantly larger number of themes in the middle phase of recovery. However, this does not necessarily imply a “longer” phase. Furthermore, a number of themes were common in multiple phases of recovery. For example, spirituality is a common theme of Phase 1 and Phase 3 (Young & Ensing, 1999). Improvement in item hierarchy, thus the measure’s ability to identify people with very low or high recovery, might be achieved by adding items that are more exclusive to the initial phase or the last phase. The analysis of reliability and validity in Rasch framework did not lead the researcher to delete any of the 30 items.

**Limitations of the Study**

As discussed before, the study had two methodological limitations. One limitation of the study was related to sampling and consequent limitations regarding the generalizability of the results. The data was collected from 19 agencies which might have had different group referral procedures. This might have affected the sampling of participants. Moreover, participation in these groups is voluntary. However, this limitation is less problematic in Rasch framework as compared to traditional methods due to separate estimations for item and person parameters. Another limitation of the study was the self-report format of the measure. Choosing Rasch methodology addressed this
limitation to some extent as Rasch analysis enables the researcher to identify inconsistencies in responses.

Another limitation of the study was the missing demographical information. As mentioned before 18% of the participants reported ethnicity information when completing pre-treatment data. For the gender, the percentage of valid data was 25.5%. This result in large was due to the fact that Ohio Mental Health Consumer Outcomes System-Adult Consumer Form (Ohio Department of Mental Health, 2000) that was used to collect demographical information in the WMR program was discontinued while during the data collection process. There was no indication that the distribution of those with no ethnicity and gender information were different than the rest of the group. However, the large percentage of missing demographic information might have elevated the percentage of people who identified their ethnicity as Native American/Pacific Islander or Latino’s. Furthermore, the demographic data suggested that Asian population was not represented at all although they constitute 1.7% of Ohio residents, according to United States Census Bureau (2011). This could, in part, attributed to this group’s underrepresentation in mental health system due to unique barriers they face (e.g. language, stigma about mental illness) (Strickland, 2010).

Implications and Recommendations

The MHRM measure was found to have acceptable reliability. In other words, these results are replicable with a different sample or a different set of items. The items measure a unidimensional construct that is on a low/high continuum. However, the analyses revealed the measure cannot adequately differentiate those at the higher and lower levels of recovery from those who have abilities closer to the average range. In
order to evaluate recovery, and thus treatment outcomes, the measure’s ability to differentiate people with very high or very low abilities needs to be improved. This would be a challenging task given the personal nature of recovery and the reality that the progress of recovery does not happen in linear and mutually exclusive phases (Deegan, 1988). However, realizing this task would be invaluable for treatment planning and treatment evaluation. The MHRM is used in the WMR program as well as other places within and outside the USA to measure impact of treatments on recovery. The measure’s ability to differentiate persons with different levels of recovery is essential for accurate evaluation of change over-time with and without treatment. It is recommended that the broader goals of the current study continue to be pursued. Specifically, adding items or revising existing items in order to target persons at the initial and later phases of recovery is recommended. Several strategies are suggested to aid the development of new items and revising existing one. One strategy is considering the relevance and prognosis of particular items as a function of stage of recovery. As mentioned before, Andresen et al. (2010) studied the earlier version of the MHRM with 41-items (this version did not include all of the 30 items in the current measure), and concluded that for empowerment and overcoming stuckness domains, change across stages of recovery was a function of the associated stage of recovery. These results highlight the importance of targeting the whole range of recovery with specific items for the particular range. Once new items or revised items are developed to target the high and low ends of recovery, future research should investigate the possibility of using different item sets for different ranges. This is suggested by different results found in the literature regarding how the particular components of recovery measures correlate with other outcome measures. For example,
Stewart and Kopache (2002) concluded that “righteous anger,” a component of recovery, was positively correlated with symptom distress. On the other hand, in their study that used Empowerment scale in addition to other measures, Bullock et al. (2000) concluded that righteous anger (a component of the Empowerment scale) increased with treatment as measured by the difference between pre-treatment and post-treatment scores. It is possible that people at earlier stages of recovery might experience righteous anger different than those who are at a later stage of recovery. Thus an item that measures righteous anger might be measuring different experiences as a function of the stage. Although such a possibility is in part addressed by Rasch model fit, it remains to be a guideline in revising items and adding new ones for better targeting the whole range of recovery.

Furthermore, in order to target those at a lower level of recovery, addition of items that assess the suffering aspect of recovery could be considered. In a review of recovery measures, Ralph et al. (2000) pointed out that the items in contemporary recovery measures concentrated on the positive side of recovery and that in order for the measurements to truly mirror the range of recovery, they should include anguish and similar aspects of recovery.

Furthermore, it is recommended that adding new items and revision of existing items are done in the framework of emerging themes and approaches in mental health field. The last century witnessed dramatic changes in the field of healthcare (National Research Council, 2009). In December 2011, SAMSHA identified a new definition of recovery, and emphasized that it was still a “working definition.” This working definition
of recovery included four major dimensions of recovery: health, home (a stable and safe place to live), purpose in life, and community (relationships and social networks).

The dimension of “health” in this definition was reflective of the emerging theme of integrated health care and more general the concept of mind-body connection. The idea that physical and mental health is inseparable and thus should be assessed, treated, and studied together has been changing the way consumers and providers conceptualize recovery (Schultz et al., 2009; SAMSHA, 2011). New treatment models such as “medical home” reflect this trend and aim to provide integrated health services (Schultz et al., 2009). Similarly, the recent “health home” model in Ohio aims to provide integrated health services to individuals with severe and persistent mental illness (Ohio Department of Mental Health, 2012). One of strategic initiatives and action plans by SAMHSA (“Recovery Support”) emphasized the need for integrated services because behavioral problems frequently co-occur with, and may exacerbate, other medical conditions. Indeed, research revealed that individuals with serious mental illnesses and co-occurring disorders die at age 53, on average, due to treatable medical conditions caused by modifiable risk factors, including smoking, obesity, high blood pressure, and substance abuse (Parks, et al., 2006). Currently, the only MHRM item that is directly related to the physical wellness is item 13, (“I eat nutritious meals every day.”). Potential new items might reflect the low and high levels of physical health and wellness.

Another emerging theme is patient-centered care. Patient-centered behavioral health care is defined as involving consumers in all aspects of behavioral health services, including planning, policy development, training, delivery, administration, and research (SAMSHA, 2011). In addition to continuing to stress the concept of empowerment,
patient-centered care emphasizes the importance of communication between providers and consumers (Schultz et al., 2009; SAMSHA, 2011). This concept naturally has altered the providers’ and consumers’ standards and expectations, in other words, “markers” of recovery. For example, medication compliance was previously suggested as one of the markers of recovery. More contemporary markers of recovery regarding psychopharmacological treatment might include effective communication with health care providers that goes beyond consumers’ behavioral compliance with psychopharmacological prescriptions by the healthcare provider (Schultz et al., 2009).

Furthermore, as seen in SAMSHA’s new definition of recovery, another prominent theme of recovery is improving all components of quality of life (SAMSHA, 2011). Young and Ensing (1999) had identified improved life standards as one of the markers for the later (last) phase of recovery. Currently, MHRM items 28, (“I engage in work or other activities that enrich myself and the world around me”) and 30 (“I have enough money to spend on extra things or activities that enrich my life,”) target two major components of satisfaction with one’s life standards. SAMSHA has broadened the components of life standard under “purpose” domain to include any activity that create a sense of contribution, including education, volunteerism, family caretaking, creative endeavors, independence, and resources to participate in society. Our understanding of what a good life standard is (in the context of recovery) has been changing in the last decade parallel to advances in the treatment of mental illness in addition to the societal and cultural changes. As treatments advance and the stigma about mental illness decreases, individuals with mental illness increasingly aim for a life that is fulfilling based on their values. This transformation in the field has also been impacted by the
biopsychosocial approach to mental illness. This approach acknowledges the role of psychosocial environment in prognosis and treatment of mental illness. Linehan (1993) described this view in the context of one of the most severe and challenging mental health problems: suicidal behavior and Borderline Personality Disorder (BPD), and stated that the patient’s desire to die is rarely a result of “distorted” thinking. She argued that in fact most patients with BPD have valid reasons for wanting to be dead, in the context of their current life (e.g. interpersonal, vocational, financial problems). Therefore, she suggested that the treatment goal should not be defined as preventing suicide (although that too would be essential so that people can build a better life), but that the goal of treatment should be building a life that is worth living. New MHRM items might include those that target one’s satisfaction with their quality of life.

In summary, new items could be added to the current MHRM, taking into account recent developments and emerging best practices in the health care field to better target different levels of recovery with its contemporary understanding. Moreover, the items that target areas that are common to multiple phases of recovery might be broken into 2-3 items to reflect different levels within the concept. For example, item 27, ("I advocate for the rights of myself and others with mental health problems,") targets both the middle phase (advocating for one’s self) and the late phase (advocating for others) of recovery (Young & Ensing, 1999). Items 25, ("When I am feeling low, my religious faith or spirituality helps me feel better,") and item 26, ("My religious faith or spirituality supports my recovery,") target spirituality, which is a theme of both the initial phase and the later (i.e. last) phase of recovery. The extent to which the role of spirituality is different in each phase is a question that needs further analysis. If these roles can be
differentiated, these items could be revised to reflect spirituality as it relates to the initial phase (e.g. easy item(s)), and to the late phase (e.g. difficult item(s)).
References


Harding, C. M., Brooks, G. W., Ashikaga, T., Strauss, J. S., & et al. (1987). The Vermont longitudinal study of persons with severe mental illness: II. Long-term outcome of


Appendix A

Mental Health Recovery Measure (MHRM)©
(Young & Bullock, 2003)

Your First Name: ________________________   Last Four Numbers of Your SSN: _____________

The goal of this questionnaire is to find out how you view your own current recovery process. The mental health recovery process is complex and is different for each individual. There are no right or wrong answers. Please read each statement carefully, with regard to your own current recovery process, and indicate how much you agree or disagree with each item by filling in the appropriate circle.

SD = Strongly Disagree  D = Disagree  NS = Not Sure  A = Agree  SA = Strongly Agree

1. I work hard towards my mental health recovery. O O O O O O
2. Even though there are hard days, things are improving for me. O O O O O O
3. I ask for help when I am not feeling well. O O O O O O
4. I take risks to move forward with my recovery. O O O O O O
5. I believe in myself. O O O O O O
6. I have control over my mental health problems. O O O O O O
7. I am in control of my life. O O O O O O
8. I socialize and make friends. O O O O O O
9. Every day is a new opportunity for learning. O O O O O O
10. I still grow and change in positive ways despite my mental health problems. O O O O O O
11. Even though I may still have problems, I value myself as a person of worth. O O O O O O
12. I understand myself and have a good sense of who I am. O O O O O O
13. I eat nutritious meals everyday. O O O O O O
14. I go out and participate in enjoyable activities every week. O O O O O O
15. I make the effort to get to know other people. O O O O O O
<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>16.</td>
<td>I am comfortable with my use of prescribed medications.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>17.</td>
<td>I feel good about myself.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>18.</td>
<td>The way I think about things helps me to achieve my goals.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>19.</td>
<td>My life is pretty normal.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>20.</td>
<td>I feel at peace with myself.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>21.</td>
<td>I maintain a positive attitude for weeks at a time.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>22.</td>
<td>My quality of life will get better in the future.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>23.</td>
<td>Every day that I get up, I do something productive.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>24.</td>
<td>I am making progress towards my goals.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>25.</td>
<td>When I am feeling low, my religious faith or spirituality helps me feel better.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>26.</td>
<td>My religious faith or spirituality supports my recovery.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>27.</td>
<td>I advocate for the rights of myself and others with mental health problems.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>28.</td>
<td>I engage in work or other activities that enrich myself and the world around me.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>29.</td>
<td>I cope effectively with stigma associated with having a mental health problem.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>30.</td>
<td>I have enough money to spend on extra things or activities that enrich my life.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

Thank you for completing this measure.

The MHRM© was developed with the help of mental health consumers by researchers at the University of Toledo, Department of Psychology. This research was supported through a grant from the Ohio Department


Appendix B

Outline of the Recovery Model

I. INITIATING RECOVERY: OVERCOMING “STUCKNESS”

A. Acknowledging and accepting illness
B. Having desire and motivation to change
C. Finding/having a source of hope/inspiration
   1. Spirituality
   2. Other people/role models

II. MIDDLE PHASE: REGAINING WHAT WAS LOST AND MOVING FORWARD (458)

A. Recovery is a process of discovering and fostering self-empowerment
   1. Actively taking control and responsibility for own life
      a. Taking responsibly for own recovery
         (1) Self-monitoring & being proactive
         (2) General coping with illness
      b. Taking responsibility for behaviors
         (1) Growing up & confronting problems
         (2) Stopping destructive habits
         a. Abuse of alcohol
         b. Abuse of caffeine and other drugs
   2. Developing other empowering attitudes
      a. Believing in self/self-confidence
b. Being determined and working hard

c. Having courage to challenge self & take risks

d. Being patient with self

B. Recovery is a process of learning and self-redefinition

1. Gaining insight about self

a. Recapturing parts of old self

b. Discovering new aspects of self

2. Gaining insight about the relationship between self and illness

a. Learning that there is more to self than illness

b. Learning / realizing that illness is not my fault

3. Gaining insight about living in the world

a. Learning to live in the moment

b. Learning to be honest with self and others

C. Recovery is a process of returning to basic functioning

1. Being able to take care of self

a. Taking care of basic needs

1) Eating

2) Hygiene

3) Basic physical health

b. Medication compliance and monitoring

c. Exercising

d. Taking care of living space

e. Doing things I enjoy
2. Being active
   a. Exercising
   b. Doing things I enjoy
   c. Vocational activities
   d. Participating in programs
   e. Following through with ideas
3. Connecting with others
   a. Consumer friends
   b. Family
   c. Professionals
   d. Community

III. LATER PHASE: IMPROVING QUALITY OF LIFE

A. Recovery involves striving to attain an overall sense of well-being
   1. Increasing self-esteem
   2. Feeling at peace/calm
   3. Feeling “normal”/stable
   4. Adopting more functional thinking
   5. Caring about things vs. apathy

B. Recovery involves striving to reach new potentials of higher functioning
   1. Finding meaning, purpose in life
      a. Spirituality
      b. Helping other consumers
c. Vocational activities
d. Using creative abilities

2. Ability to improve standard of living
a. Can afford certain possessions
   1. Nice Home
   2. Car
b. Can afford leisure activities
   3. Increased independence/self-sufficiency
   4. Marinating positive focus
   5. Symptom reduction

Appendix C

Item Statistics

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>TOTAL</th>
<th>TOTAL</th>
<th>MODEL</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>PT-MEASURE</th>
<th>EXACT MATCH</th>
<th>ESTIM</th>
<th>ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER</td>
<td>SCORE</td>
<td>COUNT</td>
<td>MEASURE</td>
<td>S.E.</td>
<td>MNSQ</td>
<td>ZSTD</td>
<td>MNSQ</td>
<td>ZSTD</td>
<td>CORR.</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>------------</td>
<td>-------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>3491</td>
<td>1114</td>
<td>-.81</td>
<td>.04</td>
<td>1.28</td>
<td>5.1</td>
<td>1.39</td>
<td>6.2</td>
<td>.44</td>
</tr>
<tr>
<td>10</td>
<td>3214</td>
<td>1113</td>
<td>-.41</td>
<td>.04</td>
<td>1.71</td>
<td>-7.1</td>
<td>.73</td>
<td>-5.8</td>
<td>.61</td>
</tr>
<tr>
<td>11</td>
<td>3161</td>
<td>1110</td>
<td>-.36</td>
<td>.04</td>
<td>1.78</td>
<td>-5.1</td>
<td>.72</td>
<td>-6.1</td>
<td>.61</td>
</tr>
<tr>
<td>12</td>
<td>2991</td>
<td>1106</td>
<td>-.16</td>
<td>.03</td>
<td>1.75</td>
<td>-6.2</td>
<td>.77</td>
<td>-5.0</td>
<td>.63</td>
</tr>
<tr>
<td>13</td>
<td>2523</td>
<td>1114</td>
<td>.37</td>
<td>.03</td>
<td>1.28</td>
<td>6.6</td>
<td>1.45</td>
<td>9.2</td>
<td>.51</td>
</tr>
<tr>
<td>14</td>
<td>2432</td>
<td>1112</td>
<td>.45</td>
<td>.03</td>
<td>1.13</td>
<td>3.3</td>
<td>1.28</td>
<td>6.1</td>
<td>.58</td>
</tr>
<tr>
<td>15</td>
<td>2770</td>
<td>1114</td>
<td>.11</td>
<td>.03</td>
<td>1.03</td>
<td>8.10</td>
<td>1.08</td>
<td>1.7</td>
<td>.57</td>
</tr>
<tr>
<td>16</td>
<td>3075</td>
<td>1093</td>
<td>-.31</td>
<td>.04</td>
<td>1.51</td>
<td>9.8</td>
<td>1.87</td>
<td>9.2</td>
<td>.41</td>
</tr>
<tr>
<td>17</td>
<td>2768</td>
<td>1092</td>
<td>.05</td>
<td>.03</td>
<td>1.66</td>
<td>-9.4</td>
<td>.65</td>
<td>-8.6</td>
<td>.69</td>
</tr>
<tr>
<td>18</td>
<td>2894</td>
<td>1092</td>
<td>-.09</td>
<td>.03</td>
<td>1.67</td>
<td>-8.7</td>
<td>.67</td>
<td>-7.7</td>
<td>.65</td>
</tr>
<tr>
<td>19</td>
<td>2302</td>
<td>1095</td>
<td>.55</td>
<td>.03</td>
<td>1.94</td>
<td>-1.4</td>
<td>1.04</td>
<td>1.0</td>
<td>.63</td>
</tr>
<tr>
<td>2</td>
<td>3172</td>
<td>1113</td>
<td>-.36</td>
<td>.04</td>
<td>1.92</td>
<td>-1.8</td>
<td>1.20</td>
<td>3.8</td>
<td>.54</td>
</tr>
<tr>
<td>20</td>
<td>2498</td>
<td>1091</td>
<td>.34</td>
<td>.03</td>
<td>1.76</td>
<td>-6.4</td>
<td>.77</td>
<td>-5.8</td>
<td>.68</td>
</tr>
<tr>
<td>21</td>
<td>2365</td>
<td>1089</td>
<td>.47</td>
<td>.03</td>
<td>1.86</td>
<td>-3.8</td>
<td>.91</td>
<td>-2.2</td>
<td>.66</td>
</tr>
<tr>
<td>22</td>
<td>3169</td>
<td>1094</td>
<td>-.43</td>
<td>.04</td>
<td>1.78</td>
<td>-5.2</td>
<td>.85</td>
<td>-2.9</td>
<td>.59</td>
</tr>
<tr>
<td>23</td>
<td>2709</td>
<td>1091</td>
<td>.11</td>
<td>.03</td>
<td>1.88</td>
<td>-3.0</td>
<td>.92</td>
<td>-1.8</td>
<td>.61</td>
</tr>
<tr>
<td>24</td>
<td>2890</td>
<td>1092</td>
<td>-.09</td>
<td>.03</td>
<td>1.68</td>
<td>-8.3</td>
<td>.71</td>
<td>-6.7</td>
<td>.64</td>
</tr>
<tr>
<td>25</td>
<td>2928</td>
<td>1094</td>
<td>-.12</td>
<td>.03</td>
<td>1.33</td>
<td>6.8</td>
<td>1.46</td>
<td>8.4</td>
<td>.50</td>
</tr>
<tr>
<td>26</td>
<td>2973</td>
<td>1090</td>
<td>-.19</td>
<td>.03</td>
<td>1.30</td>
<td>6.3</td>
<td>1.47</td>
<td>8.4</td>
<td>.49</td>
</tr>
<tr>
<td>27</td>
<td>2930</td>
<td>1090</td>
<td>-.14</td>
<td>.03</td>
<td>1.17</td>
<td>3.8</td>
<td>1.26</td>
<td>5.0</td>
<td>.49</td>
</tr>
<tr>
<td>28</td>
<td>2700</td>
<td>1094</td>
<td>.13</td>
<td>.03</td>
<td>1.89</td>
<td>-2.6</td>
<td>.92</td>
<td>-1.9</td>
<td>.61</td>
</tr>
<tr>
<td>29</td>
<td>2581</td>
<td>1092</td>
<td>.25</td>
<td>.03</td>
<td>1.93</td>
<td>-1.8</td>
<td>1.02</td>
<td>4.1</td>
<td>.59</td>
</tr>
<tr>
<td>3</td>
<td>2940</td>
<td>1111</td>
<td>-.09</td>
<td>.03</td>
<td>1.22</td>
<td>4.8</td>
<td>1.23</td>
<td>4.6</td>
<td>.51</td>
</tr>
<tr>
<td>30</td>
<td>1787</td>
<td>1092</td>
<td>1.06</td>
<td>.03</td>
<td>1.70</td>
<td>9.9</td>
<td>2.02</td>
<td>9.9</td>
<td>.50</td>
</tr>
<tr>
<td>4</td>
<td>2837</td>
<td>1115</td>
<td>.04</td>
<td>.03</td>
<td>1.28</td>
<td>6.2</td>
<td>1.67</td>
<td>9.9</td>
<td>.47</td>
</tr>
<tr>
<td>5</td>
<td>3070</td>
<td>1109</td>
<td>-.25</td>
<td>.03</td>
<td>1.84</td>
<td>-3.8</td>
<td>.92</td>
<td>-3.8</td>
<td>.63</td>
</tr>
<tr>
<td>6</td>
<td>2547</td>
<td>1109</td>
<td>-.33</td>
<td>.03</td>
<td>1.94</td>
<td>-4.1</td>
<td>1.10</td>
<td>2.3</td>
<td>.59</td>
</tr>
<tr>
<td>7</td>
<td>2680</td>
<td>1113</td>
<td>.21</td>
<td>.03</td>
<td>1.88</td>
<td>-3.2</td>
<td>.98</td>
<td>-.5</td>
<td>.62</td>
</tr>
<tr>
<td>8</td>
<td>2691</td>
<td>1110</td>
<td>.18</td>
<td>.03</td>
<td>1.10</td>
<td>2.3</td>
<td>1.22</td>
<td>4.5</td>
<td>.57</td>
</tr>
<tr>
<td>9</td>
<td>3496</td>
<td>1110</td>
<td>-.84</td>
<td>.04</td>
<td>.97</td>
<td>-.5</td>
<td>.90</td>
<td>-1.8</td>
<td>.50</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>------------</td>
<td>-------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>MEAN</td>
<td>2819.5</td>
<td>1101.8</td>
<td>.00</td>
<td>.03</td>
<td>1.01</td>
<td>-5.1</td>
<td>1.10</td>
<td>1.0</td>
<td>.47</td>
</tr>
<tr>
<td>S.D.</td>
<td>353.0</td>
<td>10.0</td>
<td>.39</td>
<td>.00</td>
<td>.26</td>
<td>5.5</td>
<td>.35</td>
<td>5.8</td>
<td>.72</td>
</tr>
</tbody>
</table>
Appendix D

Scale X Category Diagnostics

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>OBSERVED</th>
<th>OBSVD SAMPLE</th>
<th>INFIT OUTFIT</th>
<th>ANDRICH</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABEL</td>
<td>SCORE</td>
<td>COUNT</td>
<td>%</td>
<td>AVRGE EXPECT</td>
<td>MNSQ</td>
</tr>
<tr>
<td>0</td>
<td>2513</td>
<td>9</td>
<td>-0.88</td>
<td>-1.06</td>
<td>1.22</td>
</tr>
<tr>
<td>1</td>
<td>4377</td>
<td>16</td>
<td>-0.32</td>
<td>-0.16</td>
<td>.81</td>
</tr>
<tr>
<td>2</td>
<td>12411</td>
<td>46</td>
<td>.74</td>
<td>.75</td>
<td>.81</td>
</tr>
<tr>
<td>3</td>
<td>7734</td>
<td>29</td>
<td>2.27</td>
<td>2.23</td>
<td>1.04</td>
</tr>
</tbody>
</table>

MISSING 6394 19

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

CATEGORY PROBABILITIES: MODES - Structure measures at intersections

P
R 1.0 +
O |
B | 3
A | 00 3333|
B | .8 + 00 33 +
I | 00 33 |
L | 0 222222 33 |
T | .6 + 0 22 22 33 |
Y | 0 22 22 22 33 |
O | 0 2 3322 |
F | .4 + 0 2 3 22 |
R | 111 111 111 33 22 |
E | 11 22 0 11 3 22 |
S | .2 + 111 2 00 111 33 22 |
P | 111 22 0 33*1 2222 |
O | 222 00*33 111 2 |
N | 22222 33333 00000 11111 |
S | .0 333333333333333 |
E | -3 -2 -1 0 1 2 3 4 |

PERSON [MINUS] ITEM MEASURE
## Appendix E

### Scale Z Category Diagnostics

**SUMMARY OF CATEGORY STRUCTURE. Model="R"
---
-------------------------------------------------------------
<table>
<thead>
<tr>
<th>CATEGORY LABEL</th>
<th>OBSERVED SCORE</th>
<th>OBSERVED SAMPLE</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>ANDRIC THRESHOLD</th>
<th>CATEGORY MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-.90</td>
<td>6890 21</td>
<td>1.09</td>
<td>1.32</td>
<td>NONE</td>
<td>( -2.11)</td>
</tr>
<tr>
<td>1</td>
<td>-.35</td>
<td>6019 18</td>
<td>.93</td>
<td>1.02</td>
<td>-.51</td>
<td>-.76</td>
</tr>
<tr>
<td>2</td>
<td>.22</td>
<td>12411 38</td>
<td>.94</td>
<td>.88</td>
<td>-.77</td>
<td>.52</td>
</tr>
<tr>
<td>3</td>
<td>1.44</td>
<td>7734 23</td>
<td>.97</td>
<td>1.00</td>
<td>1.28</td>
<td>1.28 ( 2.48)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>390 1</td>
<td>.22</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
-------------------------------------------------------------------
OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

**CATEGORY PROBABILITIES: MODES - Structure measures at intersections**

```
P 1.0 +
| O |
| B .8 +
| A 0000
| 0000
| T 00 333
| I 00 33 |
| L 0 33 |
| Y 00 2222 33 |
| O 00 22 3*2 |
| F 0 2 33 222 |
| R 022 33 22 |
| E 1111 33 222 |
| S .2 +
| P 11111 22 01111 33 |
| O 1111111 |
| N 2222222 33333 00000 1111111 |
| S .0 +
| E -3 -2 -1 0 1 2 3 |
```

PERSON [MINUS] ITEM MEASURE
### Category Diagnostics (87 Misfitting People)

#### SUMMARY OF CATEGORY STRUCTURE

Model = "R"

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>OBSERVED</th>
<th>OBSVD SAMPLE</th>
<th>INFIT OUTFIT</th>
<th>ANDRICH</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABEL</td>
<td>SCORE</td>
<td>COUNT %</td>
<td>AVRGE EXPECT</td>
<td>MNSQ</td>
<td>MNSQ</td>
</tr>
<tr>
<td>0   0</td>
<td>777 34</td>
<td>-.36  -.39</td>
<td>1.06  1.09</td>
<td>NONE</td>
<td>( -1.17)</td>
</tr>
<tr>
<td>1   1</td>
<td>239 11</td>
<td>.00   .21</td>
<td>.81   .79</td>
<td>1.07</td>
<td>.00</td>
</tr>
<tr>
<td>2   2</td>
<td>1246 55</td>
<td>.99   .96</td>
<td>.97   .99</td>
<td>-1.07</td>
<td>( 1.17)</td>
</tr>
<tr>
<td>MISSING</td>
<td>348 13</td>
<td>-.19</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

---

**CATEGORY PROBABILITIES: MODES**

Structure measures at intersections

```
| P | 1.0 | +                   |
| R | 0.0 | 000000 222222 +    |
| B | 000 0000 222222 +   |
| A | 0000 222222 +       |
| I | 00 22 00 22       +  |
| L | 00 22 00 22 +      |
| T | 00 22 00 22 +      |
| Y | 00 22 00 22 +      |
| O | 00 22 00 22 +      |
| F | 00 22 00 22 +      |
| R | 22 00 22 00 +      |
| E | 22 00 22 00 +      |
| S | 22 00 22 00 +      |
| P | 1111111111111111110 000 + |
| O | 111111111111111111111111111111111 000 + |
| N | 111111111111111111111111111111111 000 + |
| S | 111111111111111111111111111111111 000 + |
| E | 111111111111111111111111111111111 000 + |
```

PERSON [MINUS] ITEM MEASURE

-2 -1 0 1 2
## Category Diagnostics (986 Fitting People)

**SUMMARY OF CATEGORY STRUCTURE. Model=”R”**

<table>
<thead>
<tr>
<th>CATEGORY LABEL</th>
<th>OBSERVED</th>
<th>OBSVD SAMPLE COUNT</th>
<th>INFIT MNSQ</th>
<th>OUTFIT MNSQ</th>
<th>ANDRICH THRESHOLD</th>
<th>CATEGORY MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>5934</td>
<td>-1.90</td>
<td>-1.88</td>
<td>1.02</td>
<td>1.05</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>12040</td>
<td>-0.02</td>
<td>-0.04</td>
<td>0.91</td>
<td>0.89</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>5719</td>
<td>2.01</td>
<td>2.03</td>
<td>1.04</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MISS</td>
<td>-0.72</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

**CATEGORY PROBABILITIES: MODES - Structure measures at intersections**

```
    P  R  1.0 +
     |   |    +
     O |
     B |
     A |
     B .8 + | 0 |
     I | 000   111   222 |
    L | 00   111111  111111  22 |
    I | 00   111  111  22 |
    T .6 + | 00   111   111  22 |
        | 00   11 |
        | 11   00 |
        | 11   00 |
    Y .5 + | 0*1   1*2 |
    O | 11   00 |
    F .4 + | 11   11 |
        | 111  00 |
        | 111  00 |
    R | 11   000 |
    E | 111  00 |
    S  .2 + | 000   222 |
          | 00*22 |
    P | 22222222 |
    O | 00000000 |
    N | 22222222 |
    S  .0 + | 00000000 |
    E  -3 -2 -1 0 1 2 3 |

PERSON [MINUS] ITEM MEASURE
```
Appendix H

Item Hierarchy (87 Misfitting People)
Appendix I

Item Hierarchy (986 Fitting People)

MEASURE PERSON - MAP --ITEM
-<more>|<rare>

6 . + |
5 . + |
4 . + |
3 .# |
2 .# |
1 .### MHRM15 MHRM23 MHRM28 MHRM8
0 ###### MHRM17 MHRM4
-1 ##### MHRM12 MHRM18 MHRM24 MHRM3
-2 ###### MHRM25 MHRM27
-3 ###### MHRM10 MHRM11 MHRM2 MHRM26 MHRM5
-4 ###### S MHRM16
-5 .## |
-6 .# |

EACH "#" IS "\.", EACH "." IS "\060"
Appendix J

Category Diagnostics (170 Misfitting People)

SUMMARY OF CATEGORY STRUCTURE. Model="R"

<table>
<thead>
<tr>
<th>CATEGORY LABEL</th>
<th>OBSERVED COUNT</th>
<th>OBSERVED %</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>ANDRICH</th>
<th>CATEGORY MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1297</td>
<td>31</td>
<td>1.04</td>
<td>1.19</td>
<td>-1.29</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>671</td>
<td>16</td>
<td>.81</td>
<td>.83</td>
<td>.58</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>2223</td>
<td>53</td>
<td>1.05</td>
<td>1.03</td>
<td>-.58</td>
<td>1.29</td>
</tr>
<tr>
<td>MISSING</td>
<td>908</td>
<td>18</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

CATEGORY PROBABILITIES: MODES - Structure measures at intersections

PERSON [MINUS] ITEM MEASURE
Appendix K
Item Hierarchy (170 Misfitting People)
Appendix L

Category Diagnostics (932 People)

### SUMMARY OF CATEGORY STRUCTURE

- **Model**: "R"
- **Category Structure**: 

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>OBSERVED</th>
<th>OBSVD SAMPLE</th>
<th>INFIT OUTFIT</th>
<th>ANDRICH</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LABEL</td>
<td>SCORE COUNT</td>
<td>%AVRGE EXPECT</td>
<td>MNSQ</td>
<td>MNSQ</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>5593</td>
<td>-2.08</td>
<td>-2.04</td>
<td>0.99</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>11740</td>
<td>-0.03</td>
<td>-0.07</td>
<td>0.92</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>5107</td>
<td>2.09</td>
<td>2.13</td>
<td>1.06</td>
</tr>
</tbody>
</table>

**MISSING**: 5460 20  -.83

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

### CATEGORY PROBABILITIES: MODES - Structure measures at intersections

```
P  +------------------------------------------+
R  1.0 +
O  |
B  |
A  |
B  .8 +
I  | 000 11111111111111    222|
L  | 00 11111111111111    22 |
I  | 00 11111111111111    22 |
T  .6 +
| 00 11 11 11 22 +
Y  | 00 11 11 11 22 +
O  | 00 11 11 00 22 11 |
F  .4 +
| 11 00 22 11 +
R  | 11 00 22 11 |
E  | 111 00 222 111 |
S  .2 +
| 111 00 22 |
P  | 00 222 |
O  | 000 000 |
N  | 22222222 0000000 |
S  0 +22222222222222222 |
E  | 00000000000000000000+|
```
## Appendix M

### Final Item Hierarchy (932 People)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Person - Map - Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>&lt;more&gt;</td>
</tr>
<tr>
<td>5</td>
<td>.# +</td>
</tr>
<tr>
<td></td>
<td>.#</td>
</tr>
<tr>
<td></td>
<td>.#</td>
</tr>
<tr>
<td></td>
<td>.#</td>
</tr>
<tr>
<td></td>
<td>.#</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>.#</td>
</tr>
<tr>
<td></td>
<td>.#</td>
</tr>
<tr>
<td></td>
<td>.#</td>
</tr>
<tr>
<td>3</td>
<td>.# T+</td>
</tr>
<tr>
<td></td>
<td>.#</td>
</tr>
<tr>
<td></td>
<td>.#</td>
</tr>
<tr>
<td></td>
<td>.#</td>
</tr>
<tr>
<td></td>
<td>.#</td>
</tr>
<tr>
<td>2</td>
<td>.## + MHRM30</td>
</tr>
<tr>
<td></td>
<td>.##</td>
</tr>
<tr>
<td></td>
<td>.##</td>
</tr>
<tr>
<td></td>
<td>.##</td>
</tr>
<tr>
<td></td>
<td>.##</td>
</tr>
<tr>
<td></td>
<td>.##</td>
</tr>
<tr>
<td>1</td>
<td>.### + MHRM19</td>
</tr>
<tr>
<td></td>
<td>.###</td>
</tr>
<tr>
<td></td>
<td>.###</td>
</tr>
<tr>
<td></td>
<td>.###</td>
</tr>
<tr>
<td></td>
<td>.###</td>
</tr>
<tr>
<td></td>
<td>.###</td>
</tr>
<tr>
<td></td>
<td>.###</td>
</tr>
<tr>
<td></td>
<td>.###</td>
</tr>
<tr>
<td></td>
<td>.###</td>
</tr>
<tr>
<td>0</td>
<td>.#####</td>
</tr>
<tr>
<td></td>
<td>.#####</td>
</tr>
<tr>
<td></td>
<td>.#####</td>
</tr>
<tr>
<td></td>
<td>.#####</td>
</tr>
<tr>
<td></td>
<td>.#####</td>
</tr>
<tr>
<td></td>
<td>.#####</td>
</tr>
<tr>
<td></td>
<td>.#####</td>
</tr>
<tr>
<td></td>
<td>.#####</td>
</tr>
<tr>
<td>-1</td>
<td>.###### +</td>
</tr>
<tr>
<td></td>
<td>.######</td>
</tr>
<tr>
<td></td>
<td>.######</td>
</tr>
<tr>
<td></td>
<td>.######</td>
</tr>
<tr>
<td></td>
<td>.######</td>
</tr>
<tr>
<td>-2</td>
<td>.##### +</td>
</tr>
<tr>
<td></td>
<td>.#####</td>
</tr>
<tr>
<td></td>
<td>.#####</td>
</tr>
<tr>
<td></td>
<td>.#####</td>
</tr>
<tr>
<td>-3</td>
<td>- +</td>
</tr>
<tr>
<td></td>
<td>.# T</td>
</tr>
<tr>
<td></td>
<td>.#</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
<tr>
<td>-4</td>
<td>- +</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
<tr>
<td>-5</td>
<td>-</td>
</tr>
</tbody>
</table>

Each "#" is "." "Each "." is "." "O 6"
Appendix N

Item Fit statistics

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>MEASURE</th>
<th>COUNT</th>
<th>MODLSE</th>
<th>IN.MSQ</th>
<th>IN.ZST</th>
<th>OUT.MS</th>
<th>OUT.ZS</th>
<th>PTMEAS</th>
<th>WEIGHT</th>
<th>OBSMA</th>
<th>EXPMA</th>
<th>DISCRM</th>
<th>PME-E</th>
<th>NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.66</td>
<td>813.0</td>
<td>.07</td>
<td>1.18</td>
<td>3.55</td>
<td>1.16</td>
<td>1.82</td>
<td>.52</td>
<td>1.00</td>
<td>69.0</td>
<td>69.3</td>
<td>.77</td>
<td>.62</td>
<td>MHRM1</td>
</tr>
<tr>
<td>2</td>
<td>-.72</td>
<td>776.0</td>
<td>.07</td>
<td>.89</td>
<td>-2.18</td>
<td>.89</td>
<td>-1.72</td>
<td>.66</td>
<td>1.00</td>
<td>73.0</td>
<td>69.2</td>
<td>1.11</td>
<td>.65</td>
<td>MHRM2</td>
</tr>
<tr>
<td>3</td>
<td>-.21</td>
<td>816.0</td>
<td>.07</td>
<td>1.25</td>
<td>4.80</td>
<td>1.26</td>
<td>4.16</td>
<td>.58</td>
<td>1.00</td>
<td>63.8</td>
<td>68.8</td>
<td>.70</td>
<td>.67</td>
<td>MHRM3</td>
</tr>
<tr>
<td>4</td>
<td>.07</td>
<td>743.0</td>
<td>.07</td>
<td>1.35</td>
<td>6.19</td>
<td>1.34</td>
<td>5.29</td>
<td>.56</td>
<td>1.00</td>
<td>63.8</td>
<td>68.6</td>
<td>.59</td>
<td>.67</td>
<td>MHRM4</td>
</tr>
<tr>
<td>5</td>
<td>-.63</td>
<td>743.0</td>
<td>.08</td>
<td>.87</td>
<td>-2.71</td>
<td>.88</td>
<td>-1.90</td>
<td>.74</td>
<td>1.00</td>
<td>71.7</td>
<td>69.4</td>
<td>1.17</td>
<td>.67</td>
<td>MHRM5</td>
</tr>
<tr>
<td>6</td>
<td>.56</td>
<td>648.0</td>
<td>.08</td>
<td>1.09</td>
<td>1.66</td>
<td>1.07</td>
<td>1.05</td>
<td>.67</td>
<td>1.00</td>
<td>69.8</td>
<td>69.0</td>
<td>.90</td>
<td>.68</td>
<td>MHRM6</td>
</tr>
<tr>
<td>7</td>
<td>.32</td>
<td>720.0</td>
<td>.08</td>
<td>.93</td>
<td>-1.40</td>
<td>.88</td>
<td>-2.12</td>
<td>.72</td>
<td>1.00</td>
<td>71.4</td>
<td>68.7</td>
<td>1.10</td>
<td>.68</td>
<td>MHRM7</td>
</tr>
<tr>
<td>8</td>
<td>.22</td>
<td>790.0</td>
<td>.07</td>
<td>.95</td>
<td>-1.08</td>
<td>.95</td>
<td>-.88</td>
<td>.72</td>
<td>1.00</td>
<td>68.9</td>
<td>68.7</td>
<td>1.08</td>
<td>.67</td>
<td>MHRM8</td>
</tr>
<tr>
<td>9</td>
<td>-1.44</td>
<td>828.0</td>
<td>.07</td>
<td>.96</td>
<td>-.86</td>
<td>.94</td>
<td>-.73</td>
<td>.59</td>
<td>1.00</td>
<td>73.5</td>
<td>69.4</td>
<td>1.03</td>
<td>.63</td>
<td>MHRM9</td>
</tr>
<tr>
<td>10</td>
<td>-.66</td>
<td>767.0</td>
<td>.07</td>
<td>.64</td>
<td>-8.04</td>
<td>.64</td>
<td>-6.42</td>
<td>.74</td>
<td>1.00</td>
<td>80.8</td>
<td>69.5</td>
<td>1.39</td>
<td>.66</td>
<td>MHRM10</td>
</tr>
<tr>
<td>11</td>
<td>-.59</td>
<td>790.0</td>
<td>.07</td>
<td>.80</td>
<td>-4.38</td>
<td>.79</td>
<td>-3.69</td>
<td>.71</td>
<td>1.00</td>
<td>75.1</td>
<td>69.3</td>
<td>1.23</td>
<td>.66</td>
<td>MHRM11</td>
</tr>
<tr>
<td>12</td>
<td>-.22</td>
<td>717.0</td>
<td>.08</td>
<td>.76</td>
<td>-4.90</td>
<td>.75</td>
<td>-4.40</td>
<td>.74</td>
<td>1.00</td>
<td>75.6</td>
<td>69.1</td>
<td>1.27</td>
<td>.67</td>
<td>MHRM12</td>
</tr>
<tr>
<td>13</td>
<td>.90</td>
<td>785.0</td>
<td>.07</td>
<td>1.43</td>
<td>8.05</td>
<td>1.47</td>
<td>7.05</td>
<td>.55</td>
<td>1.00</td>
<td>59.9</td>
<td>68.5</td>
<td>.45</td>
<td>.67</td>
<td>MHRM13</td>
</tr>
<tr>
<td>14</td>
<td>.87</td>
<td>799.0</td>
<td>.07</td>
<td>1.07</td>
<td>1.40</td>
<td>1.03</td>
<td>.61</td>
<td>.68</td>
<td>1.00</td>
<td>67.9</td>
<td>68.7</td>
<td>.93</td>
<td>.67</td>
<td>MHRM14</td>
</tr>
<tr>
<td>15</td>
<td>.31</td>
<td>805.0</td>
<td>.07</td>
<td>.89</td>
<td>-2.27</td>
<td>.87</td>
<td>-2.39</td>
<td>.70</td>
<td>1.00</td>
<td>71.8</td>
<td>68.6</td>
<td>1.13</td>
<td>.67</td>
<td>MHRM15</td>
</tr>
<tr>
<td>16</td>
<td>-.78</td>
<td>783.0</td>
<td>.07</td>
<td>1.41</td>
<td>7.43</td>
<td>1.44</td>
<td>5.98</td>
<td>.50</td>
<td>1.00</td>
<td>65.0</td>
<td>69.1</td>
<td>.50</td>
<td>.66</td>
<td>MHRM16</td>
</tr>
<tr>
<td>17</td>
<td>.08</td>
<td>714.0</td>
<td>.08</td>
<td>.70</td>
<td>-6.34</td>
<td>.68</td>
<td>-5.81</td>
<td>.79</td>
<td>1.00</td>
<td>78.2</td>
<td>69.1</td>
<td>1.36</td>
<td>.69</td>
<td>MHRM17</td>
</tr>
<tr>
<td>18</td>
<td>-.09</td>
<td>685.0</td>
<td>.08</td>
<td>.65</td>
<td>-7.32</td>
<td>.66</td>
<td>-6.08</td>
<td>.76</td>
<td>1.00</td>
<td>78.4</td>
<td>69.3</td>
<td>1.39</td>
<td>.68</td>
<td>MHRM18</td>
</tr>
<tr>
<td>19</td>
<td>1.24</td>
<td>740.0</td>
<td>.08</td>
<td>1.01</td>
<td>.21</td>
<td>.97</td>
<td>-.39</td>
<td>.69</td>
<td>1.00</td>
<td>74.6</td>
<td>69.3</td>
<td>1.00</td>
<td>.67</td>
<td>MHRM19</td>
</tr>
<tr>
<td>20</td>
<td>.75</td>
<td>720.0</td>
<td>.08</td>
<td>.81</td>
<td>-4.16</td>
<td>.75</td>
<td>-4.41</td>
<td>.77</td>
<td>1.00</td>
<td>76.3</td>
<td>68.9</td>
<td>1.27</td>
<td>.68</td>
<td>MHRM20</td>
</tr>
<tr>
<td>21</td>
<td>1.01</td>
<td>737.0</td>
<td>.08</td>
<td>.92</td>
<td>-1.76</td>
<td>.85</td>
<td>-2.54</td>
<td>.75</td>
<td>1.00</td>
<td>74.8</td>
<td>69.0</td>
<td>1.14</td>
<td>.68</td>
<td>MHRM21</td>
</tr>
<tr>
<td>22</td>
<td>-.97</td>
<td>694.0</td>
<td>.08</td>
<td>.77</td>
<td>-4.63</td>
<td>.76</td>
<td>-3.49</td>
<td>.67</td>
<td>1.00</td>
<td>78.1</td>
<td>69.8</td>
<td>1.25</td>
<td>.64</td>
<td>MHRM22</td>
</tr>
<tr>
<td>23</td>
<td>.33</td>
<td>751.0</td>
<td>.07</td>
<td>.86</td>
<td>-2.99</td>
<td>.83</td>
<td>-3.19</td>
<td>.73</td>
<td>1.00</td>
<td>71.4</td>
<td>68.6</td>
<td>1.18</td>
<td>.67</td>
<td>MHRM23</td>
</tr>
<tr>
<td>24</td>
<td>-.10</td>
<td>705.0</td>
<td>.08</td>
<td>.60</td>
<td>-8.88</td>
<td>.59</td>
<td>-7.86</td>
<td>.76</td>
<td>1.00</td>
<td>80.9</td>
<td>68.8</td>
<td>1.45</td>
<td>.67</td>
<td>MHRM24</td>
</tr>
<tr>
<td>25</td>
<td>-.50</td>
<td>768.0</td>
<td>.07</td>
<td>1.21</td>
<td>3.92</td>
<td>1.19</td>
<td>2.94</td>
<td>.63</td>
<td>1.00</td>
<td>65.3</td>
<td>69.0</td>
<td>.76</td>
<td>.67</td>
<td>MHRM25</td>
</tr>
<tr>
<td>26</td>
<td>-.73</td>
<td>751.0</td>
<td>.08</td>
<td>1.11</td>
<td>2.13</td>
<td>1.10</td>
<td>1.53</td>
<td>.64</td>
<td>1.00</td>
<td>65.3</td>
<td>69.3</td>
<td>.87</td>
<td>.67</td>
<td>MHRM26</td>
</tr>
<tr>
<td>27</td>
<td>-.45</td>
<td>693.0</td>
<td>.08</td>
<td>1.27</td>
<td>4.75</td>
<td>1.27</td>
<td>3.85</td>
<td>.56</td>
<td>1.00</td>
<td>64.1</td>
<td>69.2</td>
<td>1.68</td>
<td>.67</td>
<td>MHRM27</td>
</tr>
<tr>
<td>28</td>
<td>.24</td>
<td>706.0</td>
<td>.08</td>
<td>.85</td>
<td>-3.03</td>
<td>.83</td>
<td>-3.01</td>
<td>.73</td>
<td>1.00</td>
<td>72.7</td>
<td>68.6</td>
<td>1.19</td>
<td>.68</td>
<td>MHRM28</td>
</tr>
<tr>
<td>29</td>
<td>.61</td>
<td>659.0</td>
<td>.08</td>
<td>1.02</td>
<td>.41</td>
<td>1.08</td>
<td>1.17</td>
<td>.67</td>
<td>1.00</td>
<td>71.7</td>
<td>68.9</td>
<td>.96</td>
<td>.68</td>
<td>MHRM29</td>
</tr>
<tr>
<td>30</td>
<td>2.22</td>
<td>794.0</td>
<td>.08</td>
<td>1.60</td>
<td>9.90</td>
<td>1.49</td>
<td>4.40</td>
<td>.50</td>
<td>1.00</td>
<td>61.0</td>
<td>74.3</td>
<td>.35</td>
<td>.64</td>
<td>MHRM30</td>
</tr>
</tbody>
</table>