

Psychometric Properties of an Assessment for Mental Health Recovery Programs

Dennis G. Fisher · David Pilon · Scott L. Hershberger ·
Grace L. Reynolds · Stephen C. LaMaster · Monica Davis

Received: 3 January 2007 / Accepted: 9 June 2009 / Published online: 7 July 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract The concept of recovery can be operationalized from either the point of view of the consumer, or from the perspective of the agency providing services. The Milestones of Recovery Scale (MORS) was created to capture aspects of recovery from the agency perspective. Evidence establishing the psychometric properties of the MORS was obtained in three efforts: Inter-rater reliability using staff at The Village, a multi-service organization serving the homeless mentally ill in Long Beach, California; inter-rater reliability was also obtained from Vinfen Corporation, a large provider of housing services to mentally ill persons in Boston, Massachusetts. A test–retest reliability study was conducted using staff rating of clients at The Village, and evidence for validity was obtained using the Level of Care Utilization System (LOCUS) as a validity measure. The intra-class correlation coefficient for the inter-rater reliability study was $r = .85$ (CI .81, .89) for The Village and $r = .86$ (CI .80, .90) for Vinfen Corporation; test–retest reliability was $r = .85$ (CI .81, .87); and validity coefficients for the LOCUS were at or above $r = .49$ for all

subscales except one. There is sufficient evidence for the reliability and validity of the MORS.

Keywords Recovery · Psychometric · Scale · Reliability

Background

Organizations that provide services to persons with severe and persistent mental illness are under increasing pressure to demonstrate that those services are achieving recognizable and measurable outcomes with respect to client functioning. The concept of “recovery” can be operationalized from either the perspective of the consumer of mental health services, or from the perspective of the agency providing those services, and current efforts to measure client outcomes have used both approaches (Kidd et al. 2004; Test et al. 2005).

When considered from the perspective of the consumer, recovery is assessed from the internal experience of the individuals who use terms such as “becoming empowered,” “taking charge of their own lives,” or “becoming responsible for themselves.” Other aspects to recovery include the mitigation of psychiatric symptoms and improvement in overall functioning, as well as identifying and taking on meaningful roles in life (McGlynn 1996; Uehara et al. 2003).

Several scales have been created for use with mental health consumers in community settings. These include the Psycho-Social Well-Being Scale (PSWS) (O’Hare et al. 2003), the Camberwell Assessment of Need and Behavior and Symptom Identification Scale (Trauer and Tobias 2004), and the Satisfaction With Life Scale (Test et al. 2005), which were developed to assess consumers with

D. G. Fisher (✉) · S. L. Hershberger · G. L. Reynolds
Center for Behavioral Research and Services, California State
University, Long Beach, 1090 Atlantic Avenue, Long Beach, CA
90813, USA
e-mail: dfisher@csulb.edu

G. L. Reynolds
e-mail: greynol2@csulb.edu

D. Pilon · M. Davis
Mental Health America of Los Angeles, 100 W. Broadway
#5010, Los Angeles, CA 90802, USA

S. C. LaMaster
Vinfen Corporation, 950 Cambridge St.,
Cambridge, MA 02141, USA

respect to social and other functioning, rather than to simply focus on psychiatric symptoms.

When considered from the perspective of the agency, the focus has been on several aspects. There has been a need to evaluate the level of functioning of the client, as well as the type of services used that contributed to improved functional outcomes. The Global Assessment of Functioning (GAF) is one instrument originally developed for this purpose (Endicott et al. 1976; Greenberg and Rosenheck 2005; Hall 1995; Moos et al. 2002). The Cornell Service Index is a measure of health services usage among clients of outpatient mental health services (Sirey et al. 2005), while the Clinical Strategies Implementation Scale is frequently used to assess whether evidence-based practices are used in mental health services for persons with schizophrenia (Falloon et al. 2005; Resnick 2005).

The Level of Care Utilization System (LOCUS) (Sowers et al. 1999) was developed to provide a way to assess services needs of adults and to quantify services based on the amount and scope of resources available to clients at each level of service. The LOCUS is used to assess clients on six dimensions: (1) Risk of Harm; (2) Functional Status; (3) Medical, Addictive and Psychiatric Co-Morbidity; (4) Recovery Environment; (5) Treatment and Recovery History; and (6) Engagement. Five of the dimensions are scored on a five-point scale, with one denoting minimal risk of harm, for example (dimension 1-Risk of Harm), or minimal impairment (dimension 2-Functional Status); higher numbers mean greater risk or greater impairment. One dimension (4—Recovery Environment) has two defined subscales; the first subscale denotes the Level of Stress in the Recovery Environment and the second subscale denotes the Level of Support in the Recovery Environment. As with the other dimensions, lower scores indicate a low stress environment or a highly supportive environment and higher scores represent the negative ends of the continuum.

In California in 1997, the California Association of Social Rehabilitation Agencies (CASRA) convened a group of 50 administrators, clinicians, and consumers for the purpose of creating a system to classify consumers according to their needs. The CASRA project concluded that consumers could be assigned to clusters based on their level of risk, their level of coping skills and supports, and their level of engagement with the mental health system. The movement of consumers from one group or cluster to another could itself be viewed as an outcome, could reasonably be seen as a description of “the process of recovery,” and could be counted as such by service providers.

In 2005, voters of the state of California passed the Mental Health Services Act (MHSA), which provided for steady funding of mental health services based on an additional tax. Recovery was seen as the basis for services delivered under this Act.

This focus on recovery has significant implications not only for the types of mental health services offered and the manner in which they are delivered, but also for the way in which the effectiveness (outcomes) of mental health programs and systems are evaluated. McGlynn (1996) described five major domains of outcome measurement for mental health programs. These include clinical status (how a disorder is defined, particularly in terms of the presence and severity of symptoms); functional status (the ability of an individual to perform age appropriate activities); and quality of life (the importance of different decrements in functioning on an individual’s perception of his or her quality of life); adverse events (negative outcomes such as hospitalization, mortality, incarceration that result from system problems that could be avoided with appropriate care) and satisfaction with care (the consumer’s perception of the quality of the care that she or he received). The concept of overall recovery from a disabling mental illness as a domain of outcome measurement is now of major importance for the evaluation of mental health services (McGlynn 1996; Sowers 2005).

The groups or clusters created by the CASRA workgroup provided the framework for the development of the Milestones of Recovery Scale (MORS) (Pilon et al. 2006). The MORS consists of three underlying dimensions of the consumer’s (a) level of risk, (b) level of engagement with the mental health system, and (c) level of skills and supports. The consumer’s level of risk is comprised of three primary factors: (a) the consumer’s likelihood of causing physical harm to self or others, (b) the consumer’s level of participation in risky or unsafe behaviors, and (c) the consumer’s level of co-occurring disorders. The consumer’s level of engagement is the degree of connection between the consumer and the mental health service system. Finally, the consumer’s level of skills and supports should be viewed as the combination of the consumer’s abilities and support network(s) and the level to which the consumer needs staff support to meet his/her needs. It should include an assessment of their skills in independent living (e.g., grooming, hygiene), cognitive impairments, whether or not they are engaged in meaningful roles in their life (e.g., school, work), and whether they have a support network of family and friends. The eight levels of the MORS can be found in Table 1.

This paper reports on the psychometric properties of the MORS.

Methods

Inter-Rater Reliability Study: Long Beach, California

The inter-rater reliability study took place at The Village Integrated Service Agency during the month of October

Table 1 The Milestones of Recovery (MORS) and how it is used

Milestone	Risk	Engagement skills and support	
1. Extreme risk	5	N/A	0
2. High risk/not engaged	4	0	1
3. High risk/engaged	4	1	1
4. Poorly coping/not engaged	3	0	2
5. Poorly coping/engaged	3	1	2
6. Coping/rehabilitating	2	0/1	3
7. Early recovery	1	0/1	4
8. Advanced recovery	0	0	5

2005. All active clients were rated by two to five raters. All clients were rated by a psychiatrist or a case manager and either a neighborhood leader or one other staff person who knew them well. There were a total of 49 raters who rated 431 clients. The intra-class correlation coefficient was calculated using PROC MIXED (Littel et al. 2006) in SAS version 9.1.3. PROC MIXED provides the within and between rater variance components required for the calculation of the intra-class correlation even when the number of raters for each client differs. For determination of acceptable inter-rater reliability, .70 was chosen as an acceptable level (Nunnally and Bernstein 1994).

Inter-Rater Reliability Study: Vinfen Corporation, Boston, Massachusetts

Vinfen Corporation is the largest provider of non-profit housing service to people with psychiatric disabilities in New England. Vinfen's housing program includes 80 programs in the Psychiatric Rehabilitation Division (PRD). These include homeless outreach, residential services, supported housing, the Program of Assertive Community Treatment (PACT), with specialized programs for transitional aged youth, people living with HIV and AIDS, and people with co-occurring mental health and substance abuse disabilities.

In November 2005, one of the co-authors on this paper provided 2 days of consultation to Vinfen Corporation on the use of the MORS. This included 3 h of training to managers and staff of four pilot programs. In 2–3 h increments, each program assigned initial ratings to all persons served. A range of program types across the PRD was included in the pilot ($N = 105$). Vinfen has a total of 240 possible slots for clients, including a PACT team with 72 persons served; an outreach team serving 114 persons; a transitional aged youth program serving 19 persons; and supportive housing serving 35 persons. Each program rated all participants by consensus at the same time each month. Three programs rated participants all "in one sitting." One

program assigned ratings to 25% of the participants per week for each month.

After training the staff, the actual pilot study took place in April 2006. Each client was assigned a primary rater (usually a case manager) and a secondary rater. Each rater was blind to the other raters' rating. A total of 105 clients were rated by two individuals and both ratings on each client took place on the same day.

Test–Retest Reliability Study

The test–retest reliability study was conducted at two points in time during the month of September 2005 at The Village in Long Beach, California. Three hundred and eighty-one clients were rated at both points in time (431 at time 1 and 381 at time 2). The time interval between ratings ranged from 10–20 days.

Validity Study

A score on an existing measure, the Level of Care Utilization System (LOCUS) (Sowers et al. 1999, 2003) was obtained on all clients on whom a MORS score was obtained for 6 months, January through June 2005. The LOCUS and the MORS were obtained on all clients for each of these 6 months. Spearman correlation coefficients were obtained and .49 was used as meeting acceptable validity (Nunnally and Bernstein 1994).

Results

The demographics of the clients on whom the inter-reliability data were obtained can be found in Table 2. The inter-rater reliability achieved using clients and staff of The Village Integrated Service Center was $r = .85$ (95% CI = .81, .89). The inter-rater reliability using clients and staff of Vinfen Corporation was $r = .86$, (95% CI = .80, .90). The test–retest reliability using clients and staff of The Village was $r = .85$ (95% CI .81, .87).

Clients from each of the Vinfen program participated in the inter-rater reliability study. This included 57 from the PACT program, 16 Transitional Aged Youth (TAY), 16 from the outreach program, and 17 from the supported housing program. There were some differences between Vinfen clients who participated in the inter-rater reliability study and those who did not. While there were no significant differences between the rated and unrated clients based on gender, there was a significant difference found on race/ethnicity ($\chi^2(3) = 9.71$, $P = .02$) with greater numbers of White clients and fewer Black and Asian clients participating in the inter-rater reliability study. Clients who were rated were younger ($M = 45.13$, $SD = 13.58$)

Table 2 Demographics

Variable	The village (<i>N</i> = 381)	Vinfen corporation (<i>N</i> = 105)
Age	<i>M</i> = 37.28 (<i>SD</i> = 10.93)	<i>M</i> = 45.13 (<i>SD</i> = 13.58)
Gender		
Male	198 (52%)	44 (55%) ^a
Female	183 (48%)	36 (45%)
Race/ethnicity		
White	192 (50%)	55 (69%) ^a
Black	131 (34%)	20 (25%)
Asian	4 (1%)	1 (1%)
Hispanic	38 (10%)	0
Native American	5 (1%)	0
Other	11 (3%)	4 (8%)
Global assessment		
Functioning score	Not available	<i>M</i> = 53.63 (<i>SD</i> = 7.71)

^a Does not sum to 105 due to missing data

compared to clients who were not rated (*M* = 48.99, *SD* = 10.78, $t(174) = 2.06$, $P = .04$). Clients who were rated had higher GAF scores (*M* = 53.63, *SD* = 7.71) compared to clients who were not rated (*M* = 48.07, *SD* = 8.32, $t(174) = 3.25$, $P = .0017$).

The validity coefficients and 95% confidence intervals using the LOCUS can be found in Table 3. Validity at the acceptable $r = .49$ level was achieved for all LOCUS subscales except the Level of Support subscale.

Discussion

This paper reported on the reliability and validity of a measure for assessing client outcomes in the persistently severely mentally ill. Inter-rater and test–retest reliabilities were good overall; two independent sites were used to obtain inter-rater reliability, providing a strong test of the instrument's ability to be used consistently among several raters, in different settings, and with different populations. The values of the obtained reliabilities were remarkably similar.

The test–retest reliability of the MORS was also good and reached acceptable levels.

One measure of validity, the LOCUS, was used in this study. The LOCUS was subjected to rigorous reliability and validity studies when it was developed (Sowers et al. 1999; Sowers et al. 2003). It provides exceptional validation of the MORS with respect to convergent construct validity. One subscale of the LOCUS, the Level of Support, did not meet the stated criterion of .49. The Medical, Addictive, and Psychiatric Co-Morbidity (MAP) subscale met the criterion of .49, however, the 95% confidence interval fell below .49. We do not feel that this is a problem with the MORS in that all instruments will have stronger abilities to discriminate in some areas than in others.

In the multiple studies of inter-rater, test–retest, and validity reported here, the MORS was subjected to methodologically rigorous evaluation. Compared to the reliability and validity techniques used for other instruments that have relied on scoring of vignettes or clinician case studies (Sowers et al. 1999), we feel our methodology is strong.

It is important to note that the MORS is designed to be used as an administrative tool, not a clinical one. Its purpose is to describe the general parameters of recovery, not to prescribe the individual process of recovery. No classification system can do justice to the uniqueness of the individuals that it attempts to classify. We propose that the MORS should be used for purposes of program accountability and the establishment of benchmarks for programs; it should guide the collaboration between staff and the consumers in the development and refinement of treatment plans, as well as provide a framework for staff within which they can think and act on the current status of program participants.

There are several limitations to this study that must be noted. Firstly, the time period between the two inter-rater reliability ratings was different at the two sites. Both of the ratings at the Vinfen site in Massachusetts took place on the same day, while the inter-rater reliability ratings at the Village in Long Beach took place a week apart. These are different intervals and we cannot assess what impact this difference in timing may have had on the ratings. Secondly, there are a limited number of variables on both of the samples. No LOCUS measures were obtained on the Vinfen clients and the Global Assessment of Functioning was not obtained on the Village clients. Therefore, we have different scales on the two samples, reducing our ability to compare differences between the two samples. Finally,

Table 3 MORS overall rating and LOCUS subscale validity coefficients and 95% confidence intervals

	Risk of harm	Functional status	MAP	Level of stress	Level of support	Treatment recovery history	Engagement
MORS	.70	.72	.44	.50	.41	.70	.67
95% CI	.70, .87	.71, .90	.44, .47	.50, .55	.41, .44	.70, .87	.66, .80

clients from the Village were similar to clients at Vinfen who received services under the Program of Assertive Community Treatment (PACT) model; however, Vinfen also had clients from two additional modalities of treatment, the Transitional Aged Youth (TAY) and outreach clients. These TAY and outreach clients may differ from the other Vinfen clients and the Village clients in important ways that we have not measured.

Some future directions in the adaptation and use of the MORS include consideration of the following questions. First, are different services more or less effective at different milestones of recovery? What is the “typical” path of a person in recovery? Can such a path be described by use of the MORS? Can we hold service providers accountable for moving people through the milestones? And finally, should we set expectation for service providers to move certain percentages of their consumers to higher milestones over a set amount of time?

Acknowledgments This research was supported in part by contract number MH050002 from the Department of Mental Health, Los Angeles County, California.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry*, *33*(6), 766–771.
- Falloon, I. R. H., Economou, M., Palli, A., Malm, U., Mizuno, M., Murakami, M., et al. (2005). The clinical strategies implementation scale to measure implementation of treatment in mental health services. *Psychiatric Services*, *56*(12), 1584–1590.
- Greenberg, G. A., & Rosenheck, R. A. (2005). Using the GAF as a national mental health outcome measure in the Department of Veterans Affairs. *Psychiatric Services*, *56*(4), 420–426.
- Hall, R. C. (1995). Global assessment of functioning. A modified scale. *Psychosomatics*, *36*, 267–275.
- Kidd, S. A., Styron, T. H., Carlson, S., & Hoge, M. (2004). Utilization management for residential programs. *Administration and Policy in Mental Health*, *31*(6), 435–448.
- Littel, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS system for mixed models* (2nd ed.). Cary, NC: SAS Institute, Inc.
- McGlynn, E. A. (1996). Setting the context for measuring patient outcomes. *New Directions for Mental Health Services*, *71*, 19–32.
- Moos, R. H., Nichol, A. C., & Moos, B. S. (2002). Global assessment of functioning ratings and the allocation and outcomes of mental health services. *Psychiatric Services*, *53*(6), 730–737.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory*. New York, NY: McGraw-Hill, Inc.
- O’Hare, T., Sherrer, M. V., Connery, H. S., Thornton, J., LaButti, A., & Emrick, K. (2003). Further validation of the psycho-social well-being Scale (PSWS) with community clients. *Community Mental Health Journal*, *39*(2), 115–129.
- Pilon, D., LaMaster, S., & Fisher, D. G. (2006). *Knowing it when we see it: A proposed system for measuring recovery*. Paper presented at the United States Psychiatric Rehabilitation Association Conference, Phoenix, AZ.
- Resnick, S. G. (2005). Fidelity measurement, the CSI, and the current evidence base. *Psychiatric Services*, *56*(12), 1592.
- Sirey, J. A., Meyers, B. S., Teresi, J. A., Bruce, M. L., Ramirez, M., Raue, P. J., et al. (2005). The cornell service index as a measure of health service use. *Psychiatric Services*, *56*(12), 1564–1569.
- Sowers, W. (2005). Transforming systems of care: The American Association of Community Psychiatrists guidelines for recovery oriented services. *Community Mental Health Journal*, *41*(6), 757–774.
- Sowers, W., George, C., & Thompson, K. (1999). Level of care utilization system for psychiatric and addiction services (LOCUS): A preliminary assessment of reliability and validity. *Community Mental Health Journal*, *35*(6), 545–563.
- Sowers, W., Pumariega, A., Huffine, C., & Fallon, T. (2003). Level-of-care decision making in behavioral health services: The Locus and the Calocus. *Psychiatric Services*, *54*(11), 1461–1463.
- Test, M. A., Greenberg, J. S., Long, J. D., Brekke, J. S., & Burke, S. S. (2005). Construct validity of a measure of subjective satisfaction with life of adults with serious mental illness. *Psychiatric Services*, *56*(3), 292–300.
- Trauer, T., & Tobias, G. (2004). The Camberwell assessment of need and behaviour and symptom identification scale as routine outcome measures in a psychiatric disability rehabilitation and support service. *Community Mental Health Journal*, *40*(3), 211–221.
- Uehara, E. S., Srebnik, D. S., & Smukler, M. (2003). Statistical and consensus-based strategies for grouping consumers in mental health level-of-care schemes. *Administration and Policy in Mental Health*, *30*(4), 287–306.