Rasch analysis of the illness management and recovery scale-clinician version

Alan B. McGuire PhD HSPP,[1,2] Jacob Kean PhD CCC-SLP,[1,3,4] Kelsey Bonfils BS,[5] Jade Presnell MA[6] and Michelle P. Salyers PhD[7,8]

[1]Research Scientist, Health Services Research and Development, Richard L. Roudebush VAMC, Indianapolis, IN, USA
[2]Assistant Scientist, Department of Psychology, Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA
[3]Assistant Research Professor, Department of Physical Medicine and Rehabilitation, Indiana University School of Medicine, Indianapolis, IN, USA
[4]Research Scientist, Regenstrief Institute, Inc. and Indiana University Center for Health Services and Outcomes Research, Indianapolis, IN, USA
[5]Doctoral Student in Clinical Psychology, Department of Psychology, Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA
[6]Clinical Care Manager, Midtown Community Mental Health Center, Indianapolis, IN, USA
[7]Associate Professor, Department of Psychology, Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA
[8]Affiliated Scientist, Regenstrief Institute, Inc. and Indiana University Center for Health Services and Outcomes Research, Indiana, IN, USA

Abstract

Rationale, aims and objectives The illness management and recovery scale-clinician version (IMRS-C) is a measure of outcomes thought to be important indicators of progress for consumers participating in illness management and recovery (IMR). Prior research has examined the psychometric properties of the IMRS-C; extant research supports certain aspects of the scale's reliability (test–retest) and validity (sensitivity to interventions). Analyses based on Rasch provide certain advantages and have not been applied to the IMRS-C.

Method This study used an archival IMRS database including responses regarding 697 participants with severe mental illness from a variety of community-based settings. Rasch analyses were utilized to determine item functioning and utility of the IMRS-C.

Results Results of Rasch analyses using the IMRS-C as one unidimensional scale were problematic. Analyses grouping items into three separate scales measuring recovery, management and biological vulnerability were more promising, but the third scale had other limitations.

Conclusions Results suggest that the items included in the IMRS-C can form two screeners, one for recovery and one for management; items regarding biological vulnerability were inadequate. The assessment could be supplemented by more refined measures of coping/self-management and recovery constructs.

Introduction

The illness management and recovery scales (IMRSs) were created in conjunction with the IMR Implementation Toolkit [1] to provide a practical measure of consumer progress while participating in IMR. IMRS items target specific elements that the model creators thought to be important indicators of illness self-management. The IMRSs can be rated by the consumer as a self-report measure, or by a clinician; both versions use the same anchors. The IMRSs have been widely used in research [2–7] and have provided the core of evidence supporting the effectiveness of illness management and recovery [8].

Several analyses have examined the psychometric properties of the IMRSs [9–12]; a key issue has been the dimensionality and internal consistency of the scales. Most recently, Sklar and colleagues [12] noted the mixed results regarding internal consistency, stating 'the weaker estimates of Cronbach's alpha reported to date suggest the items may not be strongly interrelated' (p. 222). Consequently, they conducted a confirmatory factor analysis (CFA) on data from 10 659 service users from a large U.S. county in order to attempt to resolve this issue. Unfortunately, while the CFA suggested a three-factor model fit better than a one-factor model, the authors also indicated that Cronbach's analysis supported the use of a one-factor model, leaving the issue of dimensionality, and therefore appropriate use and scoring of the scale, yet unsettled.

Rasch analysis (aka single-parameter item response analysis) comprises an increasingly used set of methods to develop or refine

measures. Rasch models address several deficiencies of traditional psychometric approaches [13], including: the sample dependence of person ability estimates; the lack of comparability of scores obtained from different sets of items measuring the same construct; the necessary completion of all items to allow comparisons between respondents; the lack of interval-level raw scores; and the concentration on group-level characteristics. In contrast, Rasch models produce sample independent ability estimates. That is, when the data fit the model, the estimation of the item difficulties is not dependent on the distribution of the person abilities in the calibration samples. Rasch models also produce interval-level measures. In other words, the distance from 1 to 2 on the scale is the same amount of increase in the construct (e.g. self-management) as the increase from 10 to 11). In addition, Rasch models validate items instead of a set of items, enabling the selection of few items that reliably differentiate ability across the latent continuum.

The current study seeks to conduct a Rasch analysis on the IMRS-clinician version (IMRS-C). By overcoming the difficulties of traditional psychometric approaches, we hoped that Rasch would provide more direct guidance on the adequacy of the items in assessing the underlying construct, and provide guidance as to whether scores can be used as an overall score (measuring one dimension) or whether multiple theoretical dimensions are reflected in the IMRS-C. Once dimensionality is clarified, we propose to use Rasch analyses to examine the functioning of items within the scale.

Methods

Sample

Participants included 697 participants who completed the IMRSs as part of one of four larger studies [6,7,14,15]. Participants were adults with severe mental illness receiving outpatient mental health treatment from participating community mental health centres in Indiana. Participants included slightly more men ($n = 366$, 52.7%) and were predominately white ($n = 570$. 83.2%); the remainder of the sample included African American ($n = 65$, 9.5%), American Indian ($n = 14$, 2.0%), Asian/Pacific Islander ($n = 8$, 1.2%), Hispanic ($n = 11$, 1.6%) or other ($n = 17$, 2.5%). Most participants were single/never married ($n = 355$, 51.5%) or divorced ($n = 238$, 36.6%), with the remainder single/living together ($n = 27$, 4.1%) or married ($n = 51$, 7.8%). Educational status was fairly low, with almost a third not having a high school diploma or General Educational Development (GED, $n = 184$, 26.8%); the remainder had a high school diploma/GED ($n = 239$, 34.8%) or at least some college ($n = 258$, 37.6%). Self-reported diagnoses included schizophrenia or schizoaffective disorder ($n = 228$, 65.3%), bipolar disorder ($n = 50$, 14.3%), major depressive disorder ($n = 39$, 11.2%), personality disorder ($n = 3$, 0.9%) and other ($n = 28$, 8.3%). Participants were served by assertive community treatment teams ($n = 375$, 53.8%) and other outpatient mental health programs ($n = 322$, 46.2%).

Measures

The IMRS-C [16] was developed specifically to measure illness self-management outcomes and is based on the stress-vulnerability

model [17,18]. The IMRS has fifteen 5-point items. Previous psychometric analyses have found adequate internal consistency ($\alpha > .70$), test–retest reliability and convergent validity [9–12].

Procedures

Participants in each study provided informed consent and the participants' primary clinicians (predominately case managers) completed the IMRS-C and returned it to the research team. All study procedures were approved by the university's Institutional Review Board.

Analysis

We first calculated the mean, range and standard deviation for the IMRS using SPSS 19.0. For the remaining analyses, missing data were imputed using the mode of the variable. Mode imputation was chosen over more sophisticated but less accessible methods to enable the computation of a total score by administrators and clinicians without access to advanced statistical software. We also recognize that imputing the item mode would result in conservative, rather than inflated ability estimates. We next used Rasch analysis (i.e. a 1-parameter logistic item response analysis) as implemented in WINSTEPS version 3.80.0. We first analysed the full IMRS scale and, as described below, analysed each component 'scale' that reflects an underlying dimension separately. For the full scale, dimensionality was assessed by analysing the residual variance with eigenvalues greater than 2.0 indicating the presence of a secondary dimension. Infit and outfit statistics were examined with values less than .6 or greater than 1.4 indicating poor fit. Item local independence was examined, with items sharing 15% common variance outside the main dimension considered to be locally dependent and therefore problematic for a single dimension.

Subsequently, the three separate scales were examined in terms of dimensionality as described above. Additionally, targeting was calculated as the difference between average domain measures and item calibrations, with zero being 'perfect targeting', .50–.99 'slight mistargeting', and ≥ 1.00 'substantial mistargeting' [19]. The number of distinct strata was calculated as $(4G + 1)/3$, where G is the separation index. Item probability curves were examined to ensure that all anchors were most probable at some point and that all items met the monotonicity assumption. Point-biserial correlations were used to examine the relationship between the item and the main dimension.

Results

Descriptive statistics

Descriptive statistics for item and scale scores can be seen in Table 1. Means ranged from 2.0 to 4.6. Several items had a restricted range due to ceiling or floor effects. The per cent of respondents at floor (i.e. responding '1') ranged across items from less than 1 to 44% and the percent at ceiling (i.e. responding '5') ranged across items from 5 to 82%.

Rasch analysis: one-scale approach

We first conducted a Rasch analysis with all items loaded onto one scale. This approach produced several problematic results.

Table 1 Descriptive statistics for illness management and recovery scales

Item		Mean	SD	Rating of 1 N (%)	Rating of 5 N (%)
	Overall mean	3.19	0.56	–	–
1	Progress towards goals	2.73	1.06	74 (14.8%)	23 (4.6%)
2	Knowledge	2.97	1.07	51 (10.2%)	29 (5.8%)
3	Family/friend involvement in treatment	2.88	1.26	92 (18.4%)	48 (9.6%)
4	Contact with people outside of family	3.31	1.24	27 (5.4%)	124 (24.8%)
5	Time in structured roles	1.95	1.23	268 (53.6%)	24 (4.8%)
6	Symptom distress	2.67	0.94	61 (12.2%)	6 (1.2%)
7	Impairment of functioning	2.60	0.95	69 (13.8%)	2 (0.4%)
8	Relapse prevention planning	2.58	1.12	93 (18.6%)	16 (3.2%)
9	Relapse of symptoms	3.24	1.46	84 (17.0%)	141 (28.5%)
10	Psychiatric hospitalizations	4.26	1.18	25 (5.0%)	316 (63.7%)
11	Coping	2.99	0.95	32 (6.4%)	31 (6.2%)
12	Involvement with self-help activities	2.99	1.33	56 (11.2%)	85 (17.0%)
13	Using medication effectively	4.42	0.87	11 (2.2%)	291 (58.6%)
14	Alcohol use	4.57	0.98	10 (2.9%)	274 (79.2%)
15	Drug use	4.60	0.97	11 (3.2%)	286 (82.4%)

Analyses indicated that one secondary component in the data exceeded the criterion value of 2.0. It contrasted items 6, 7, 9, 10 and 11 with the other items. This indicates these items have something in common that separate them from the assumed single dimension of IMR improvement. Similarly, several items lacked *local independence:* 2 and 8, 6 and 7, 7 and 11, 9 and 10, 14 and 15, indicating these items have more in common with each other than just the main dimension of IMR improvement. Finally, items 5, 6, 7 and 13 failed to demonstrate monotonically increasing (i.e. never decreasing) average patient ability measures across response options 1–5 on the item response structure. This occurrence is particularly concerning because it indicates the average participant answering a '3' on a given item is *lower* on IMR improvement than the participants answering a '2' on that item. We therefore concluded that a one-scale approach was inappropriate.

Rasch analyses: three-scale approach

We next conducted a series of three Rasch analyses in which we created three scales composed of the items loading on each of the three factors identified by Sklar and colleagues [12] (see Table 3). We adopted this course because items composing the second dimension in our data were the same as the second factor identified by Sklar and colleagues [12]. Moreover, pairs of items that shared common variance in our previous Rasch analysis always loaded onto the same of Sklar and colleagues' [12] factors.

Scale 1 – 'Recovery'

The first scale (assessing a 'recovery' dimension) was composed of items 1, 2, 4, 8, 12 (e.g. 'As far as Progress towards personal goals: In the past 3 months, I have come up with . . .'). Results indicated only *one dimension* was present for these items: no second dimension was identified (secondary dimension < 2.0 eigenvalue units), no items were locally dependent (residual correlations >.3), and infit and outfit statistics for items were all

within acceptable parameters (1.0 ± .4). Internal consistency was less than acceptable ($\alpha = .67$). Scale logit scores and raw scores were strongly correlated (.98).

Targeting

Difference in person means was .13, indicating the 'average' person and the 'average' item were at about the same place on the recovery dimension.

Strata

The recovery scale produced 2.19 statistically distinguishable performance levels; this means that for a given person, the scale only reliably placed the person as either 'high' or 'low' on recovery. There is too much 'noise' to make a more fine-grained determination. As with all scales, the reliability of the scale differs depending on where in the distribution a person scores; however, we can use the scale midpoint (summed scale score of 15) as an example [20]. For this point in the scale, the 95% confidence interval was ± .92 logits; therefore, a raw score >20 or better was reliably better (95% confidence interval) than the midpoint and a score <10 was reliably lower than midpoint.

Item functioning

Rasch provides several ways to examine how well each item is functioning within the model. Table 2 displays item difficulty for each item, which provides an indicator of where on the recovery dimension each item generally falls. Results indicated items generally clustered in the middle of the recovery continuum; that is, no items received high scores from even people low on the recovery continuum and no items received low scores from only respondents high on the recovery continuum. Point-measure correlations indicate the degree to which responses to a particular item correlate with the recovery dimension. Results indicated items are moderately to strongly related to the dimension. We examined item

Published 2014. This article is a U.S. Government work and is in the public domain in the USA.

385

Table 2 Item fit and point-measure correlations for the total scale

Item		Item difficulty	Count	Infit	Outfit	Point-measure correlations
1	Progress towards goals	.51	499	.96	.96	.53
2	Knowledge	.29	501	.89	.89	.57
3	Family/friend involvement in treatment	.33	499	1.36	1.41	.31
4	Contact with people outside of family	−.33	499	1.19	1.24	.41
5	Time in structured roles	1.02	500	1.12	1.19	.48
6	Symptom distress	.83	499	.93	.93	.51
7	Impairment of functioning	1.16	500	.81	.80	.62
8	Relapse prevention planning	.70	499	.84	.83	.62
9	Relapse of symptoms	−.07	495	1.00	1.00	.57
10	Psychiatric hospitalizations	−.84	496	1.14	1.33	.40
11	Coping	.15	500	.69	.69	.72
12	Involvement with self-help activities	.03	500	1.16	1.23	.47
13	Using medication effectively	−1.08	497	1.00	1.16	.40
14	Alcohol use	−1.55	346	1.05	1.03	.38
15	Drug use	−1.15	347	.94	.80	.43

Table 3 Item fit and point-measure correlations for three dimensions

Item		Item difficulty	Count	Infit	Outfit	Point–measure correlations
Recovery scale items						
1	Progress towards goals	.33	499	1.01	1.02	.63
2	Knowledge	.03	501	.87	.85	.69
4	Contact with people outside of family	−.69	499	1.28	1.34	.56
8	Relapse prevention planning	.56	499	.77	.75	.73
12	Involvement with self-help activities	−.23	500	1.08	1.05	.65
Management scale items						
6	Symptom distress	1.48	499	.94	.93	.74
7	Impairment of functioning	1.06	500	.86	.83	.77
9	Relapse of symptoms	−.53	495	.90	1.69	.77
10	Psychiatric hospitalizations	−2.00	496	1.40	1.30	.58
11	Coping	−.01	500	1.01	1.01	.73
Biology scale items						
13	Using medication effectively	.08	497	1.19	1.08	.84
14	Alcohol use	−.04	345	.93	.77	.72
15	Drug use	−.04	347	.87	.52	.72

probability curves to see that each item anchor is most probable at some point on the ability continuum; results were generally good. Only items 4 (anchor 4) and 12 (anchor 3) were never most probable. Importantly, unlike the single scale analyses, all items demonstrated monotonically increasing (i.e. never decreasing) average patient ability measures across response options 1–5 on the item response structure.

Scale 2 – 'Management'

The second scale (assessing a 'management' dimension) was composed of items 6, 7, 9, 10, 11 (e.g. 'In regards to Symptom of distress: How much do your symptoms bother you?). We included item 10 even though Sklar and colleagues [12] did not because it was consistent with the second dimension identi-

fied in our initial Rasch analysis. In addition, it made conceptual sense that hospitalization would be related to management of one's illness. Regarding dimensionality of the management scale, a potential second dimension was identified with an eigenvalue exactly on the threshold (2.0); however, no items were locally dependent. Infit statistics for items were all within acceptable parameters (± .4); however, item 9 exceeded the parameter for outfit. Internal consistency was good for this scale ($\alpha = .80$).

Targeting

Difference in person means was .1, indicating the 'average' person and the 'average' item were at about the same place on the self-management dimension.

Strata

The management scale produced 2.81 statistically distinguishable performance levels; again, the scale only reliably placed the person as either 'high' or 'low' on management. For this scale, the 95% confidence interval was ± .92 logits; therefore, a score of 18 or better was reliably better (95% confidence interval) than an average score of 15, and a score of 12 or lower was reliably lower than average.

Item functioning

As shown in Table 2, items on the management scale covered a wide range of the continuum, with some items (e.g. impairment in functioning) very difficult and others (e.g. psychiatric hospitalization) very easy. Point-measure correlations indicated moderate to strong correlations with the main dimension. Item probability curves indicated that all anchors were most probable at some point and that all items met the monotonicity assumption. In general, results indicated that items were functioning well within the model.

Scale 3 – 'Biology'

Results for the three 'biology' items (13, 14, 15; e.g. 'As far as Using Medication Effectively: How often do you take your medication as prescribed?') were less encouraging. Regarding dimensionality of this scale, no second dimension accounting for greater than 2.0 eigenvalue units was identified; however, all items were locally dependent, with correlated variance outside the model exceeding 15%. Infit/outfit statistics were acceptable, with no items with values outside of $1.0 \pm .4$.

Targeting

The difference in person means was 3.03, indicating the 'average' person was over 3 logit scores above the 'average' item on the biology dimension. In other words, participants generally endorsed the highest anchors on all items (see also Table 1 regarding ceiling effects).

Strata

Less than one (.68) strata could be identified; the biology scale was not reliable enough to categorize people into distinct categories.

Item functioning

Point-biserial correlations indicated moderate to strong correlations with the main dimension. Item probability curves indicated that for item 13, a response of '3' was never most probable. For item 14, responses 2 and 4 were most probable, but at a very small range of the continuum. For item 15, responses 2 and 4 were never most probable. All items met the monotonicity assumption.

Discussion

Our results indicate the IMRS-C is most appropriately modelled using three conceptual dimensions: recovery, management and biology. When modelling the scale as three scales, there is little advantage to the use of a logit transformation or anchor modifications – raw scores are adequately reliable. However, although three dimensions were identified, the items assessing the third dimension did not perform well enough to be considered a scale. In addition, the management and recovery scales are useful as rough screeners, but cannot be reliably used for more fine-grained measurement due to the relatively large error around ability estimates.

Our results join a growing body of literature that demonstrates the IMRS-C measures more than one construct [11,12]. The three dimensions indicated by our results are consistent with the stress vulnerability model of mental illness [21]. People strive to learn and implement strategies to manage their illnesses. In addition, other factors affect biological vulnerabilities to mental disability (i.e. substances and medication). Self-management strategies and management of biological factors work together to affect changes in community functioning, role functioning and other indicators of 'recovery'. To combine the IMRS into one total score is to ignore this conceptual distinction and lose meaningful information about the person being measured.

The IMRS was constructed with each meaningful element represented by one item. This decision was made with practicality in mind, and the resulting scale can be administered with minimal burden to busy clinical staff. The trade-off to this approach is reliability and sensitivity. Based on our analyses, the management and recovery scales derived from the IMRS-C represent acceptable and usable *screeners* for their respective constructs. For each, only about two distinct strata, (i.e. two groups of 'high' or 'low') could be identified reliably. Clinical implications from our findings indicate that the scales could be used to triage consumers into appropriate services. For instance, at intake the management scale may quickly assess people who may benefit from basic self-management education. It would be inappropriate, though, to use the management scale to determine incremental change after an intervention. The implications for research are even more striking. The IMRS-C overestimates the amount of variability or 'noise', thereby reducing power in statistical analyses. Previous research showing positive impacts on the IMRS-C has likely greatly underestimated the true effect of IMR. Subsequent research using more sensitive measures could better estimate effect sizes and would require much smaller samples to demonstrate statistical significance.

The items tapping biological vulnerability are conceptually consistent with the stress vulnerability model of mental health; however, these items did not function well as a scale in our analyses. All the items are 'too easy', reflected by the ceiling effects for the items, which have been found in all studies reporting item-level averages for the IMRS-C [9,10,12]. Three explanations could plausibly explain this ceiling effect. First, items may need to be added to capture people higher on the 'biology' continuum (i.e. people with less impairment from drugs/alcohol and who adhere more fully to medication regiments). This suggestion makes little sense, though, because a response of '5' on all three items indicates a complete lack of impairment due to alcohol/drugs and complete adherence to medication. In other words, these responses appear to be the upper conceptual limit. The second explanation is that participants in this and other psychometric studies of the IMRS-C generally do not suffer from impairment in these areas. This explanation seems implausible given the prevalence of substance use

and medication non-adherence in this population [22–26]. The final explanation is that clinicians underestimate consumers' substance use and medication non-adherence. Given that clients often under-report substance abuse or medication non-adherence [27,28], it would not be surprising if clinicians also under-report actual use. With the availability of numerous psychometrically sound measures regarding substance use (e.g. the decisional balance scale [29]; AUDIT-C [30]) and medication adherence (MARS [31]), we suggest that the biology items be omitted from the IMRS-C and established screeners be used in their place.

This study utilized a relatively large sample from various service settings to investigate the item functioning and utility of the IMRS-C. However, results are limited in that the sample was primarily white with lower levels of education. Results suggest that the items currently composing IMRS-C should be disaggregated. Two scales – management and recovery screeners – can be formed from the extant items. Biological vulnerability, although conceptually important, is not reliably measured by the intended items. The assessment could be supplemented in practice by more refined measures of coping/self-management and recovery constructs.

Acknowledgements

References

1. Gingerich, S. & Mueser, K. T. (2002) Illness Management and Recovery Implementation Resource Kit. Rockville, MD: Center for Mental Health Services, Substance Abuse and Mental Health Services Administration.

2. Bullock, W. A., O'Rourke, M., Breedlove, A., Farrer, E. & Smith, M. K. (2007) Effectiveness of the illness management and recovery program in promoting recovery: preliminary results. *New Research in Mental Health*, 17, 282–291.

3. Hasson-Ohayon, I., Roe, D. & Kravetz, S. (2007) A randomized controlled trial of the effectiveness of the illness management and recovery program. *Psychiatric Services: A Journal of the American Psychiatric Association*, 58 (11), 1461–1466.

4. Färdig, R., Lewander, T., Melin, L., Folke, F. & Fredriksson, A. (2011) A randomized controlled trial of the illness management and recovery program for persons with schizophrenia. *Psychiatric Services: A Journal of the American Psychiatric Association*, 62 (6), 606–612.

5. Levitt, A. J., Mueser, K. T., DeGenova, J., Lorenzo, J., Bradford-Watt, D., Barbosa, A., Karlin, M. & Chernick, M. (2009) Randomized controlled trial of illness management and recovery in multiple-unit

6. Salyers, M. P., Godfrey, J. L., McGuire, A. B., Gearhart, T., Rollins, A. L. & Boyle, C. (2009) Implementing the illness management and recovery program for consumers with severe mental illness. *Psychiatric Services: A Journal of the American Psychiatric Association*, 60 (4), 483–490.

7. Salyers, M. P., McGuire, A. B., Rollins, A. L., Bond, G. R., Mueser, K. T. & Macy, V. R. (2010) Integrating assertive community treatment and illness management and recovery for consumers with severe mental illness. *Community Mental Health Journal*, 46 (4), 319–329.

8. McGuire, A. B., Kukla, M., Green, A., Gilbride, D., Mueser, K. T. & Salyers, M. P. (2014) Illness management and recovery: a review of the literature. *Psychiatric Services: A Journal of the American Psychiatric Association*, 65, 171–179.

9. Salyers, M. P., Godfrey, J. L., Mueser, K. T. & Labriola, S. (2007) Measuring illness management outcomes: a psychometric study of clinician and consumer rating scales for illness self management and recovery. *Community Mental Health Journal*, 43 (5), 459–480.

10. Färdig, R., Lewander, T., Fredriksson, A. & Melin, L. (2011) Evaluation of the illness management and recovery scale in schizophrenia and schizoaffective disorder. *Schizophrenia Research*, 132, 157–164.

11. Hasson-Ohayon, I., Roe, D. & Kravetz, S. (2008) The psychometric properties of the Illness Management and Recovery scale: client and clinician versions. *Psychiatry Research*, 160 (2), 228–235.

12. Sklar, M., Sarkin, A., Gilmer, T. & Groessl, E. (2012) The psychometric properties of the Illness Management and Recovery scale in a large American public mental health system. *Psychiatry Research*, 199 (3), 220–227.

13. Shultz, K. S. & Whitney, D. J. (2005) Measurement Theory in Action: Case Studies and Exercises. Thousand Oaks, CA: Sage Publications, Inc.

14. Salyers, M. P., Stull, L., Rollins, A. & Hopper, K. (2011) The work of recovery on two assertive community treatment teams. *Administration and Policy in Mental Health and Mental Health Services Research*, 38 (3), 169–180.

15. Salyers, M. P., Matthias, M. S., Sidenbender, S. & Green, A. (2013) Patient activation in schizophrenia: insights from stories of illness and recovery. *Administration and Policy in Mental Health and Mental Health Services Research*, 40 (5), 419–427.

16. Mueser, K. T., Gingerich, S., Salyers, M. P., McGuire, A. B., Reyes, R. U. & Cunningham, H. (2004) The Illness Management and Recovery Scale (IMR). Lebanon, NH: New Hampshire-Dartmouth Psychiatric Research Center.

17. Liberman, R. P., Mueser, K. T., Wallace, C. J., Jacobs, H. E., Eckman, T. & Massel, H. K. (1986) Training skills in the psychiatrically disabled: learning coping and competence. *Schizophrenia Bulletin*, 12, 631–647.

18. Zubin, J. & Spring, B. (1977) Vulnerability: a new view of schizophrenia. *Journal of Abnormal Psychology*, 86, 103–126.

19. Duncan, P. W., Bode, R. K., Min Lai, S. & Perera, S. (2003) Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. *Archives of Physical Medicine and Rehabilitation*, 84 (7), 950–963.

20. Jette, A. M. & Haley, S. M. (2005) Contemporary measurement techniques for rehabilitation outcomes assessment. *Journal of Rehabilitation Medicine*, 37 (6), 339–345.

21. Mueser, K. T., Meyer, P. S., Penn, D. L., Clancy, R., Clancy, D. M. & Salyers, M. P. (2006) The Illness Management and Recovery program: rationale, development, and preliminary findings. *Schizophrenia Bulletin*, 32 (1), 32–43.

22. Adams, J. & Scott, J. (2000) Predicting medication adherence in severe mental disorders. *Acta Psychiatrica Scandinavica*, 101 (2), 119–124.

23. Pratt, S. I., Mueser, K. T., Driscoll, M., Wolfe, R. & Bartels, S. J. (2006) Medication nonadherence in older people with serious mental illness: prevalence and correlates. *Psychiatric Rehabilitation Journal*, 29 (4), 299–310.

24. Regier, D. A., Farmer, M. E., Rae, D. S., Lock, B. Z., Keith, S. J., Judd, L. L. & Goodwin, F. K. (1990) Comorbidity of mental disorders with alcohol and other drug abuse: results from the Epidemiologic Catchment Area (ECA) study. *Journal of the American Medical Association*, 264 (19), 2511–2518.

25. Mueser, K. T., Yarnold, P. R., Levinson, D. F., Singh, H., Bellack, A. S., Kee, K., Morrison, R. L. & Yadalam, K. G. (1990) Prevalence of substance abuse in schizophrenia: demographic and clinical correlates. *Schizophrenia Bulletin*, 16 (1), 31–56.

26. Mueser, K. T., Yarnold, P. R., Rosenberg, S. D., Swett, C. Jr, Miles, K. M. & Hill, D. (2000) Substance use disorder in hospitalized severely mentally ill psychiatric patients. *Schizophrenia Bulletin*, 26 (1), 179–192.

27. Garber, M. C., Nau, D. P., Erickson, S. R., Aikens, J. E. & Lawrence, J. B. (2004) The concordance of self-report with other measures of medication adherence: a summary of the literature. *Medical Care*, 42 (7), 649–652.

28. Johnson, T. & Fendrich, M. (2005) Modeling sources of self-report bias in a survey of drug use epidemiology. *Annals of Epidemiology*, 15 (5), 381–389.

29. Carey, K. B., Maisto, S. A., Carey, M. P. & Purnine, D. M. (2001) Measuring readiness-to-change substance misuse among psychiatric outpatients: I. Reliability and validity of self-report measures. *Journal of Studies on Alcohol*, 62 (1), 79–88.

30. Bush, K., Kivlahan, D. R., McDonell, M. B., Fihn, S. D. & Bradley, K. A. (1998) The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. *Archives of Internal Medicine*, 158 (16), 1789–1795.

31. Fialko, L., Garety, P. A., Kuipers, E., Dunn, G., Bebbington, P. E., Fowler, D. & Freeman, D. (2008) A large-scale validation study of the Medication Adherence Rating Scale (MARS). *Schizophrenia Research*, 100 (1), 53–59.