Psychometric validation of the Recovery Star

Dr Anna Good, Research Analyst, Triangle Consulting

August 2019

The psychometric properties of the Recovery Star have been examined in seven peer-reviewed journal articles (Dickens, Weleminsky, Onifade & Sugarman, 2012; Frost et al., 2017; Griffiths et al., 2015; Killaspy, White, Taylor & King, 2012; Lloyd, Williams, Machingura & Tse, 2015; McEvoy, Schauman, Mansell & Morris, 2012; Placentino, Lucchi, Scarsato & Fazzari, 2017). In addition, Triangle have conducted tests using over 4000 Recovery Star readings collected across 10 organisations (Good & Lamont, 2018). The methods used in this internal testing have been peer-reviewed in a manuscript about the Family Star Plus, which is soon to be published (Good & MacKeith, 2019). The findings from this body of research are summarised below:

Acceptability. The Recovery Star has been found to be acceptable and useful by both practitioner and service user populations (Killaspy et al., 2012; Placentino et al., 2017; Mental Health Providers Forum, 2009). In research conducted in community and inpatient mental health services, Killaspy and colleagues report that of a sample of 92 service users, 85% found it useful/very useful in 'helping them and the staff understand how they were getting on' and 'for helping them and the staff plan the support they needed'. The same percentage of staff (85%) reported that the Recovery Star was useful/very useful for care planning. The authors concluded that the Recovery Star was 'acceptable to staff and service users, with few reporting it as difficult to complete and most reporting completion within 30–60 minutes' (Killaspy et al., 2012, p.69). Another validation study reported that the Recovery Star 'was deemed acceptable by service users and keyworkers and is characterised by practical and useful visual aids' (Placentino et al., 2017, p.247). The acceptability and appropriateness of the Recovery Star when working with service users from black, Asian and minority ethnic backgrounds has also been documented, with small modifications made when creating the second edition in response to the research findings (Mental Health Providers Forum, 2009). In this cultural competency research, the majority of clients agreed that they found 'completing the Recovery Star enjoyable' and that it gave them 'a better understanding of what I want to achieve'.

Several other studies note the benefits of using the Recovery Star in keywork. For example, Esan et al. (2012) described its role in a learning disabilities service as 'opening up avenues for discussing topics covered in the domains of the Recovery Star tool which may otherwise have not been discussed as fully' (p.24). The Recovery Star is considered a 'useful tool for a collaborative approach to care planning' amongst members of the UK Routine Clinical Outcomes Measurement Network– Mental Health (MacDonald & Fugard, 2015). Finally, in an Australian context, Lloyd and colleagues (2015) concluded that it was useful in 'assisting recovery clinicians to identify areas that they needed to focus on when providing treatment and following service users' progress' (p.57).

Factor structure and internal consistency.

Identifying whether the Recovery Star has one, or more than one underlying factor is of most relevance when organisations want to aggregate readings across outcomes areas to produce a single score. A unidimensional factor structure suggests that all outcome areas represent a single underlying construct (e.g. Recovery), and that from a psychometric validation perspective at least, it is not unreasonable to produce an overall Star score. Triangle's analyses of 4000 Recovery Star readings used Parallel analysis (Horn, 1965), which is appropriate for the ordinal data produced by the Star and has been found to be superior to conventional methods for correctly identifying the number of dimensions (Hubbard & Allen, 1987; Ruscio & Roche, 2012; Zwick & Velicer, 1986). This yielded a unidimensional factor structure explaining 54% of the variance in the data (Good & Lamont, 2018), with very good internal consistency within the factor (Cronbach's α = .86). Similar findings were reported by Frost and colleagues (2017), who report a Cronbach's α of 0.88 and state that there was 'clearly a very strong and consistent first principal component in all analyses… providing strong justification for forming an overall MHRS score based on (averaging) the 10 items'. The development process and research literature supporting the outcome areas suggest that this underlying construct is related to psychological recovery.

Two studies indicated that there might instead be two clusters within the outcome areas (Dickens et al., 2012; McEvoy et al., 2012). McEvoy et al. described the two components as Emotional regulation and Social capital. However, it is worth noting that the methods used can overestimate the number of dimensions (Hubbard & Allen, 1987; Ruscio & Roche, 2012; Zwick & Velicer, 1986). Dickens and colleagues described factor 1 as representing 'internal management and personal relationships' and factor 2 representing 'external management and external relationships' and found good internal consistency within each factor. When multidimensionality is found, the subscales can still be correlated components of the same construct – e.g. Recovery, but at a service delivery level, it might be beneficial for interventions or predictive analyses to be based on specific clusters of outcome areas. In any case, from the perspective of learning from the data, producing a single score for the overall Star reading is likely to result in some loss of information concerning an individual or service user group. Triangle recommends interrogating the data by outcome area rather than producing total scores whenever possible.

Convergent validity.

Evidencing that the data from a measurement tool converges with validated tools assessing a similar construct is a valuable way to demonstrate that it is collected accurately and is measuring what it sets out to measure (Rubio, Berg-Weger, Tebb, Lee & Rauch, 2003). Placentino and colleagues (2017) found significant correlations between the Recovery Star and comparable areas of three validated measures of quality of life and level of functioning: The Health of the Nation Outcome Scales (HoNOS; Wing et al., 1998), The World Health Organisation's Quality of Life assessment (WHOQOL-BREF; WHOQOL Group, *1998b)* and The

Global Assessment of Functioning (GAF; APA, 2013).They concluded that 'most areas of the tool correlate with general quality of life and level of functioning'.

Killaspy and colleagues found convergence between the Recovery Star and the Life Skills Profile (LSP), a validated social functioning measure (Parker, Rosen, Emdur & Hazipavlov, 1991). They also examined convergence between the Recovery Star and the Mental Health Recovery Measure (MHRM; Young & Bullock, 2003), but this was described as poor convergence because the correlations between the Recovery Star and MHRM were lower than .70. However, there were in fact moderate correlations between the MHRM and all Recovery Star outcome areas except for Addiction. Addiction is not an issue for all individuals using mental health services, so readings may be at the top of the Journey of Change for many service users rendering it less well correlated with other areas at an organisational level. Nonetheless it is clearly important to include addiction in the Recovery Star, and indeed Illicit drug and alcohol abstinence is one of the four dimensions of recovery developed by the Substance Abuse and Mental Health Services Administration (SAMHSA, 2012). It may also be of relevance that Triangle's guidance for staff and service users to complete the Recovery Star collaboratively was not followed, and instead completion was service-use only for the MHRM and staff-only for the LSP. Moreover, the MHRM has been shown to have relatively weak convergence with validated measures of social functioning (van Nieuwenhuizen, Wilrycx, Moradi & Brouwers, 2014) and social functioning is central to empirically based conceptual frameworks of recovery such as the CHIME framework of Recovery Processes (Leamy, Bird, Le Boutillier, Williams, J., & Slade 2011).

Responsiveness.

Seven peer-reviewed articles have shown the responsiveness of the Recovery Star to change occurring during interactions with mental health services (Dickens et al., 2012; Frost et al., 2017; Griffiths, Heinkel & Dock, 2015; Larsen & Griffiths, 2013; McEvoy et al., 2012; Placentino et al., 2017; Raftery et al., 2017). In addition, Triangle have analysed responsiveness in a sample of more than 1000 service users who had two or more Star readings (Good & Lamont, 2018) using a Wilcoxen signed-rank test appropriate for the ordinal level of the data, and reporting effect sizes rather than just significance, in line with APA reporting standards (APA, 2010). Service users who could not move forward because they were at the top of the Journey of Change were excluded. These tests yielded a medium effect size for Managing mental health (r = .30) and small-medium effect sizes for all other areas (rs = .19 to .29).

Test-retest reliability.

Two studies have produced evidence of the temporal stability of the Recovery Star when completed within a short time period defined by the researchers (Killaspy et al., 2012; Placentino et al., 2017). In both studies the intraclass coefficient was above 0.7 for all outcome areas, indicating good test-retest reliability.

Inter-rater reliability.

The two studies reporting on test-retest reliability also examined inter-rater reliability, though each used different methods (Killaspy et al., 2012; Placentino et al., 2017). Placentino and colleagues described a clinical case to 42 workers on completion of training who then provided independent readings for each area of the Recovery Star. Reliability calculated for pairs of readings from all participants found that Cohen's kappa coefficient was greater than 0.7 in all combinations, indicating consistency in applying the Star scales. While additional clinical cases would have been desirable, this is an encouraging finding. Killaspy and colleagues instead compared staff-only readings completed within a month of each other about a service user both staff members knew well. This approach produced intraclass coefficients were between 0.46 and 0.77, with the authors describing inter-rater reliability as acceptable only for the Work area (using a .70 threshold). However, researchers commonly use a lower threshold based on guidelines reported in Fleiss (1981) and Cicchetti (1994):

Less than 0.40—poor
Between 0.40 and 0.59—Fair
Between 0.60 and 0.74—Good
Between 0.75 and 1.00—Excellent

Using these thresholds, inter-rater reliability would be considered excellent for the Work area (ICC = 0.77), good for five areas and fair for the remaining four.
In addition to this, the methodology and conclusions related to inter-rater reliability in Killaspy and colleagues' article have been critiqued in responses to the journal (Dickens & Sugarman, 2012; MacKeith, 2012) and elsewhere (e.g. Frost et al., 2017– supplementary materials). When staff complete Stars without discussion with the service user or standardised information (e.g. in a case study format), it is unsurprising that there were some discrepancies in the assigned readings.

In conclusion, there is good evidence that the Recovery Star is acceptable, useful, produces readings that converge with other relevant validated tools, is responsive to change and has temporal stability. Initial evidence suggests that workers can apply the scales consistently given the same information. Triangle is planning further research to test inter-rater reliability for the 4th Edition of the Recovery Star, and to examine its predictive validity.

References

American Psychological Association. (2010). Publication manual of the American Psychological Association (6th ed.). Washington, DC: American Psychological Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. Arlington, VA: American Psychiatric Association

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment, 6, 284–290.

Dickens, G., & Sugarman, P. (2012). Evaluating recovery star–flawed study design. *British Journal of Psychiatry, published online (http://bjp. rcpsych. org/content/201/1/65. full). Accessed*, *23.5.19*.

Dickens, G., Weleminsky, J., Onifade, Y., & Sugarman, P. (2012). Recovery Star: validating user recovery. *The Psychiatrist*, *36*(2), 45-50.

Esan, F., Case, K., Louis, J., Kirby, J., Cheshire, L., Keefe, J., & Petty, M. (2012). Implementing a patient centred recovery approach in a secure learning disabilities service. *Journal of Learning Disabilities and Offending Behaviour*, *3*, 24–35.

Fleiss JL. Statistical Methods For Rates And Proportions. Wiley; New York: 1981.

Frost, B. G., Turrell, M., Sly, K. A., Lewin, T. J., Conrad, A. M., Johnston, S., ... & Rajkumar, S. (2017). Implementation of a recovery-oriented model in a sub-acute Intermediate Stay Mental Health Unit (ISMHU). *BMC health services research*, *17*(1), 2.

Good, A., & Lamont, E. Outcomes Star™ Psychometric Factsheet: Recovery Star™ available from http://www.outcomesstar.org.uk/about-the-star/evidence-and-research/star-psychometrics/

Good, A., & MacKeith, J. (2019). Assessing family functioning: Psychometric evaluation of the Family Star Plus. Manuscript submitted for publication.

Griffiths, C. A., Heinkel, S., & Dock, B. (2015). Enhancing recovery: transition intervention service for return to the community following exit from an alternative to psychiatric inpatient admission–a residential recovery house. *The Journal of Mental Health Training, Education and Practice*, *10*(1), 39-50.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrica, 30*(2), 179–185.

Hubbard, R., & Allen, S. J. (1987). An empirical comparison of alternative methods for principal component extraction. *Journal of Business Research*, *15*(2), 173-190.

Killaspy, H., White, S. & King, M. (2012). Psychometric properties of the Mental Health Recovery Star, *British Journal of Psychiatry*, *201*, 65–70.

Leamy, M., Bird, V., Le Boutillier, C., Williams, J., & Slade, M. (2011). Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. *The British Journal of Psychiatry*, *199*, 445–452.

Lloyd, C., Williams, P. L., Machingura, T., & Tse, S. (2015). A focus on recovery: using the Mental Health Recovery Star as an outcome measure, *Advances in Mental Health*, 1–8.

Macdonald, A. J., & Fugard, A. J. (2015). Routine mental health outcome measurement in the UK, *International Review of Psychiatry*, *27*, 4, 1–14.

MacKeith, J. (2012). Square pegs and round holes: assessing the suitability of the Killaspy et al. Recovery Star validation approach. *British Journal of Psychiatry published online (http://bjp. rcpsych. org/content/201/1/65. full). Accessed*, *21*.

McEvoy, P., Schauman, O., Mansell, W., & Morris, L. (2012). The experience of recovery from the perspective of people with common mental health problems: Findings from a telephone survey. *International journal of nursing studies*, *49*(11), 1375-1382.

McNeil, B., Reeder, N. and Rich, J. (2012). *A framework of outcomes for young people*. The Young Foundation, youngfoundation.org/wp-content/uploads/2012/10/Framework-of-outcomes-for-young-people-July-2012.pdf.

Melfi, C. A., Chawla, A. J., Croghan, T. W., Hanna, M. P., Kennedy, S., & Sredl, K. (1998). The effects of adherence to antidepressant treatment guidelines on relapse and recurrence of depression. *Archives of general psychiatry*, *55*, 1128–1132.

Mental Health Providers Forum. (2009). The recovery star model and culturally competency, BAME Pilot Report. Available from www.outcomesstar.org.uk/wp.../L_The-RS-and-cultural-competency-final-report.pdf

Parker G., Rosen A., Emdur, N. & Hazipavlov, D. (1991).The Life Skills Profile: psychometric properties of a measure assessing function and disability in schizophrenia. *Acta Psychiatr Scand; 83*: 145–52.

Placentino, A., Lucchi, F., Scarsato, G., & Fazzari, G. (2017). Mental Health Recovery Star: features and validation study of the Italian version. *Rivista di psichiatria*, *52*(6), 247-254.

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social work research*, *27*(2), 94-104.

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological assessment*, *24*(2), 282.

Substance Abuse and Mental Health Services Administration. (2012). SAMHSA's Working Definition of Recovery. Rockville, MD: US Department of Health and Human Services. Retrieved from https://store.samhsa.gov/shin/content/PEP12-RECDEF/PEP12RECDEF.pdf.

van Nieuwenhuizen, C., Wilrycx, G., Moradi, M., & Brouwers, E. (2014). Psychometric evaluation of the Dutch version of the mental health recovery measure (MHRM). *International Journal of Social Psychiatry*, *60*, 162-168.

Wing, J. K., Beevor, A. S., Curtis, R. H., Park, S. G. B., Hadden, J., & Burns, A. (1998). Health of the Nation Outcome Scales (HoNOS): research and development. *The British Journal of Psychiatry*, *172*, 11-18.

WHOQOL Group (*1998b).* Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychol Med; 28*: 551–558.

Young S.L, Bullock, W.A. (2003). The Mental Health Recovery Measure. University of Toledo

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological bulletin*, *99*(3), 432.