

## Does the Outcomes Star™ accurately reflect change occurring during service provision?

The Outcomes Star is designed to enhance keywork with service users by enabling conversations about critical areas of life, informing the development of action plans and measuring changes over time. Star readings must accurately reflect changes occurring during this work. In recognition of this, we have several processes in place to assess and maximise the sensitivity and accuracy of Star readings.

In terms of Star development, new versions of the Star are co-created with those who manage and work with service users on the frontline to ensure that the detailed scales cover the key changes occurring during the provision of those services. New Stars are then piloted to ensure that the 5- or 10-point Journey of Change is sensitive to change within services and statistical tests of responsiveness are performed. Service users and keyworkers participating in the pilot are also asked to rate the extent to which the Star reflected the changes made.

Secondly, those using the Star with service users must attend training that includes how to use the scales accurately and in ways that draw out the level of engagement, progress and situations of service users. We support organisations with refresher training and encourage regular supervision meetings and auditing of Star data. We also offer free online CPD to licensed trainers.

Triangle is committed to the ongoing examination of external research and testing of published versions of the Star to ensure that baseline and follow-up readings are meaningful, and that change on the Star is reflective of what would be expected given change in other measures.

The consistency and accuracy of Star readings are assessed using tests of 'inter-rater reliability'. Whenever possible, we directly measure whether readings are assigned meaningfully by different keyworkers given the same information - this is done using the case study method reported in two of our peer-reviewed journal articles (MacKeith, 2014; Good & MacKeith, 2021). However, we recommend that the same worker stays with the person using the service throughout as developing a trusting relationship and in-depth understanding of their situation is important, especially when people are responding to multiple or complex challenges.

The extent to which Star readings reflect change can also be assessed by looking at how well Star readings and change in readings correlate with other measures as expected ('convergent validity') and whether they relate to subsequent 'hard outcomes' such as employment or offending ('predictive validity').

There is a growing body of research demonstrating convergent validity, predictive validity and inter-rater reliability of the Star. Many of these are in peer-reviewed journals authored by independent academics as well as by Triangle's Research Analyst, Dr Anna Good. We also have 'psychometric factsheets' for almost all versions of the Star. The studies are available in our [research library](#), and the most relevant are summarised below:

### a) Star readings correlate with other measures

The Star is not a pure severity measure as the initial stages capture changes in the relationship with the issue and services. The early transitions, such as beginning to acknowledge a problem, accepting help and trying things out, are often critical steps towards change in circumstances or severity but may not in themselves be accompanied by changes in the situation. This means that there will not always be a continuous relationship with more traditional severity measures. Instead, it may be that those below and above a cut-off point on the Journey of Change differ on other measures. For example, those at the final two Journey of Changes stages will often be in a better situation

than those at the first three stages. Nonetheless, studies have been able to show correlations with other validated measures:

- Placentino and colleagues (Placentino, Lucchi, Scarsato & Fazzari, 2017) found that relevant scales of the Recovery Star converged with other validated tools: the Health of the Nation Outcome Scales (Wing et al., 1998), World Health Organisation Quality of Life-BREF (Fleck, 2000) and the Global Assessment of Functioning (APA, 2013).
- Killaspy, White, Taylor and King (2012) found that Recovery Star correlated with the Life Skills Profile (Parker, Rosen, Emdur & Hazipavlov, 1991), a well-established measure of social functioning used as a routine clinical outcome tool in Australia for many years.
- Boholdt-Petersen and colleagues asked all patients admitted to a dual diagnosis ward to complete the Recovery Star alongside other validated measures. They found significant correlations in the expected directions with the Beck Depression Inventory (Beck, Steer & Brown, 1996), the Beck Anxiety Inventory (Beck & Steer, 1990) and the Robson Self-concept questionnaire (Boholdt-Petersen, Mårtensson & Bak, 2019).
- Triangle analysed Spectrum Star readings collected in a specialist Enablement service for adults with autism, finding correlations between initial Star readings, baseline costs to the service and the Rosenberg self-esteem scale (Rosenberg, 1965).
- We also found that Homelessness Star readings correlated with the New Directions Team assessment (Rinaldi, Linnell & Clenaghan, 2008) and the Warwick-Edinburgh Mental Wellbeing Scale (Tennant, Fishwick, Platt, Joseph & Stewart-Brown, 2006). The Drug and alcohol misuse scale negatively correlated with self-reported alcohol consumption.

#### **b) Star readings predict hard outcomes**

- In a report of the UK's Fulfilling lives programme, Welford et al. (2021) reported that those who had been incarcerated had lower 12-month Homelessness Star follow up readings for Self-care and living skills, Social networks and relationships and Offending. Homelessness, substance misuse and offending were also associated with significantly lower Star readings.
- Our work with a UK local authority's building resilient families and communities programme showed that service users who progressed to 7+ in the Keeping your children safe area of the Family Star Plus were much less likely to be involved in crime /antisocial behaviour (31%) than those who stayed at 1-6 on the 10-point Journey of Change (76%). Those who continued to be involved in crime/antisocial behaviour also showed a much smaller average change on the Keeping your children safe area (0.34) than those who were not involved in crime or antisocial behaviour (1.66).
- Using Family Star Plus data from another UK County Council, we found a continuous relationship between Star readings and system recorded school absence, such that the higher parents were on the Journey of Change in the Boundaries and behaviour and Education and learning areas, the lower the rate of unauthorised absence recorded at a later date. This finding is presented in our recent journal article (Good & MacKeith, 2021).

- We also reported associations between Homelessness Star readings and hard outcomes in another recently published journal article. In this research, higher Homelessness Star readings predicted accommodation and employment status as well as a greater likelihood of attending education and training. Greater change in the Managing tenancy and accommodation outcome area was also associated with being more likely to transition from homelessness to having housing (Good & MacKeith, 2021).
- In further research, Work Star readings related to employment status, with service users who ended at the higher end of the Journey of Change (7+) were significantly more likely to be employed than those below 7. For Job-search skills, two-thirds of those at the higher end of the Journey of Change were in employment compared to 44% at the lower end (Good & MacKeith, in preparation).

**c) Practitioners can accurately and consistently assign readings ('inter-rater 'reliability')**

- 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. Cohen's kappa coefficient was greater than 0.7 for all pairs of participants, indicating consistency in applying the Star scales (Placentino et al., 2017).
- Killaspy and colleagues (2012) compared staff-only readings about a service user both staff members knew. This approach produced intraclass coefficients (ICC) 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, using commonly used guidelines reported in Fleiss (1981) and Cicchetti (1994), 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 our validation article for the Family Star Plus, 65 caseworkers assigned readings to a written service user profile. Very good inter-rater reliability was found (mean Krippendorff's  $\alpha = .83$ ) when compared against readings assigned by Triangle staff (Good & MacKeith, 2021). This followed initial support for the inter-rater reliability of the Family Star reported in an earlier article (MacKeith, 2014).
- Consistency was also found to be good when the same case study methodology was used for the Carers Star, Justice Star and Well-being Star– these findings are reported in the [psychometric factsheets](#) for these Stars available on our webpage.

In conclusion, the current body of evidence suggests that Star data is meaningful and reflects changes occurring during service provision. More research is constantly being conducted to explore further and optimise the properties of the Star. For more information about research into the Star please contact Dr Anna Good ([anna@triangleconsulting.co.uk](mailto:anna@triangleconsulting.co.uk)).

## References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. Arlington, VA: American Psychiatric Association
- Beck, A. T., & Steer, R. A. (1990). *Manual for the Beck Anxiety Inventory* San Antonio, TX Psychological Corporation.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck depression inventory (BDI-II)* (Vol. 10, p. s15327752jpa6703\_13). Pearson.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardised assessment instruments in psychology. *Psychological Assessment, 6*, 284–290.
- Fleck, M. P. D. A. (2000). The World Health Organization instrument to evaluate quality of life (WHOQOL-100): characteristics and perspectives. *Ciência & Saúde Coletiva, 5*(1), 33-38.
- Fleiss JL. *Statistical Methods for Rates And Proportions*. Wiley; New York: 1981.
- Good, A., & MacKeith, J. (2021). Psychometric validation of the Homelessness Star. *Journal of Social Distress and Homelessness, 1-10*.
- Good, A., & MacKeith, J. (2021). Assessing family functioning: Psychometric evaluation of the Family Star Plus. *Family Relations, 70*(2), 529-539.
- Good, A. & MacKeith, J. (2021). Validation of the Work Star. *Manuscript in preparation*.
- Killaspy, H., White, S. & King, M. (2012). Psychometric properties of the Mental Health Recovery Star, *British Journal of Psychiatry, 201*, 65–70.
- MacKeith, J. (2014). Assessing the reliability of the Outcomes Star in research and practice. *Housing, Care and Support, 17*(4), 188-197.
- 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.
- Rinaldi, M., Linnell, D., & Clenaghan, M. (2008). The New Directions Team Assessment (Chaos Index). Retrieved from [www.meam.org.uk/wp-content/uploads/2010/05/NDTAssessment-process-summary-April-2008.pdf](http://www.meam.org.uk/wp-content/uploads/2010/05/NDTAssessment-process-summary-April-2008.pdf).
- Robson, P. (1989). Development of a new self-report questionnaire to measure self esteem. *Psychological medicine, 19*(2), 513-518.

Rosenberg, M. (1965). Rosenberg self-esteem scale (RSE). *Acceptance and commitment therapy. Measures package*, 61(52), 18.

Tennant, R., Fishwick, R., Platt, S., Joseph, S., & Stewart-Brown, S. (2006). Monitoring positive mental health in Scotland: Validating the Affectometer 2 scale and developing the Warwick-Edinburgh Mental Well-being Scale for the UK. Glasgow: NHS Health Scotland.

Welford, J., Milner, C., Moreton, R., Howe, P. Alzuabi & Roberts, J. (2021). Improving service transitions for people experiencing multiple disadvantage: Prison release – new publication. Retrieved from <https://www.fulfillinglivesevaluation.org/2021/05/>

Wing, J. K., et al. (1998). Health of the Nation Outcome Scales (HoNOS): research and development. *The British Journal of Psychiatry*, 172, 11-18.

To find out more about the Outcomes Star please see our website <https://www.outcomesstar.org.uk/> or get in touch on [info@triangleconsulting.co.uk](mailto:info@triangleconsulting.co.uk) or +44(0) 207 272 8765.