

## **The Evaluation of the Early Intervention Support Service in Northern Ireland**

### **A response from Triangle Consulting<sup>1</sup>, creators of the Outcomes Star**

In June 2018 researchers at Queen's University Belfast published their evaluation of the Early Intervention Support Service (EISS) in Northern Ireland (Winter, Neeson, Sweet & Connolly, 2018). This service uses several versions of the Outcomes Star, and the report confirms previous findings that practitioners and families are very positive about the tool. The evaluation reports a good level of change shown using the Star despite the relatively short duration of the intervention. However, we write to address serious concerns that readers will be misled by the way in which the Outcomes Star has been described:

#### **a) The use of the Star as an outcomes measure:**

Winter and colleagues use Triangle's guidance that service comparisons should be undertaken with caution as evidence that it should not be used in service evaluations. Caution in undertaking such comparisons is relevant to any outcomes measure because differences between services may reflect variation in the needs of those entering the service and the resources available to address these needs. However, the Outcomes Star is widely found to be useful in demonstrating and evidencing the impact of services (e.g. Bailey & Kerlin, 2013; Griffiths, Heinkel & Dock, 2015; York Consulting, 2013), and comparisons across services are useful provided these differences are taken into consideration, for example by comparing change for service users with a similar level of need.

The report also suggests that the data cannot be statistically analysed or collated across individuals because there is some subjectivity involved and the data is not continuous (i.e. the differences between points on the scales cannot be considered equal in the sense of continuous measures such as height). However, the authors collate Star data across service users and report tests suitable for continuous data using Star readings, despite guidance that non-parametric statistics are appropriate for this type of data.

Other measures the authors present as more objective (e.g. the self-reported Strengths and Difficulties questionnaire; Goodman, 1997) also involve the subjective judgement of the individual in rating for example 'I try to be nice to other people. I care about their feelings' and have an ordinal scale. If anything, it could be argued that the Outcomes Star encourage greater objectivity since readings involve service users and those supporting them working together to improve judgement with reference to detailed scale descriptions (which often refer to behaviour). Objectivity can be evidenced by showing that the tool predicts hard outcomes (as is shown in our Family Star Plus paper) or inter-rater reliability, which has been demonstrated for an earlier version of the Family Star Plus (MacKeith, 2014).

The authors also suggest that biases in Star readings can be inferred from their finding that a good proportion of service users in the intervention condition improved on the Star, while differences between the control and intervention groups were not statistically significant on several measures. This failure to find statistical significance is likely to be due to the small sample size, with only 33 service users in the control group, and respectable effect sizes for a number of measures considering the 12-week intervention period.

<sup>1</sup>Joy MacKeith (Director) & Dr Anna Good (Research Analyst), Triangle Consulting, Hove

**b) The psychometric evidence for the Stars:**

We disagree with the claim that the Outcomes Star has little evidence to confirm its reliability and validity. For example, the report states that the evidence base for the Recovery Star is mixed, when in fact there are at least seven peer-reviewed articles presenting positive findings for several aspects of psychometric validation (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), with only one of these reporting issues in one area (Killaspy et al., 2012) using a method that has been critiqued in responses to the journal (Dickens & Sugarman, 2012; MacKeith, 2012)

We have also conducted a psychometric validation of the Family Star Plus, which will be published later this year. Almost all versions of the Star have been validated by Dr Anna Good, Triangle's Research Analyst, using similar methods to those used in this validation paper, and there are also external peer-reviewed articles validating a number of versions of the Star (see [www.outcomesstar.org.uk/about-the-star/evidence-and-research/research-library](http://www.outcomesstar.org.uk/about-the-star/evidence-and-research/research-library) )

As part of their own psychometric testing of the Family Star Plus, the authors report finding three clusters of items within the ten outcome areas and use this as evidence that the Star lacks construct validity. There are several problems with this:

- The sample size was smaller than recommended for factor analysis
- They report that 'keyworkers only used some outcome areas and gave 10 to other areas', affecting the distribution of readings
- The statistical methods used were not designed for ordinal data (Overall, 1964) and have been shown to overestimate the number of dimensions (Hubbard & Allen, 1987; Ruscio & Roche, 2012; Zwick & Velicer, 1986).
- Having two or more sub-scales does not invalidate a measure. Indeed, they cite Dickens et al. (2012) as support for the Recovery Star, when this paper reports finding a 'valid 2-factor structure'

Our own testing found the Family Star Plus to be a unidimensional tool (Good & MacKeith, 2019). We used Parallel analysis (Horn, 1965), which is appropriate for ordinal data 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).

In conclusion, we moved to write this response in order to address the inconsistencies and misunderstandings in the EISS evaluation written by Winter and colleagues (2018). Most importantly, readers should be reassured that the Outcomes Star is a validated outcomes measurement tool, which has been shown to be useful when evaluating services.

## References

Bailey, D., & Kerlin, L. (2015). Can Health Trainers Make a Difference With Difficult-to-Engage Clients? A Multisite Case Study. *Health promotion practice*, 16, 5, 756-764.

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.

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.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, 38, 581–586

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.

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

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. doi:10.1007/BF02289447

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., Taylor, T. L., & King, M. (2012). Psychometric properties of the mental health recovery star. *The British Journal of Psychiatry*, 201(1), 65-70.

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.

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.

Mackeith, J. (2014). Assessing the reliability of the Outcomes Star in research and practice. *Housing, Care and Support*, 17(4), 188-197.

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.

Overall, J. E. (1964). Note on the scientific status of factors. *Psychological Bulletin*, 61(4), 270-276.

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.

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), 282–92.

Winter, K., Neeson, L., Sweet, D., & Connolly, P. (2018). Evaluation of the Early Intervention Support Service in Northern Ireland. Centre for Evidence and Social innovation, Queen's University Belfast.

York Consulting (2013). Family Action Family Star Evaluation: Summary Report. Retrieved 23.5.19 from [www.family-action.org.uk/content/uploads/2014/07/Family-Star-Evaluation-Summary-Report.pdf](http://www.family-action.org.uk/content/uploads/2014/07/Family-Star-Evaluation-Summary-Report.pdf)

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99(3), 432–442.