

# Report of Recovery Star Research Seminar

Hosted by the Institute of Mental Health and Triangle

**Date and location:** Centre for Mental Health, 14<sup>th</sup> June 2013

**Chair:** Professor Nick Manning, Director, Institute of Mental Health, Nottingham.

**Attendees:** Professor Chris Evans, Institute of Mental Health, Nottingham;  
Dr Sue Holttum, Senior Lecturer, Applied Psychology Canterbury Christchurch University;  
Dr Jerry Burgess, Canterbury Christchurch University;  
Sally Lawson, Head of Training and Consultancy, Community Options;  
Jo Wellsman, Peninsula Patient Involvement Group;  
Dr Richard Byng, Clinical Senior Lecturer, Peninsular College of Medicine, Universities of Exeter and Plymouth;  
Dr Mark Freestone, Queen Mary University London;  
Peter Ford, Consultant Psychologist, Head of Psychology, Community Hospitals and Care homes The Huntercombe Group;  
Joy MacKeith, co-Director Triangle;  
Sara Burns, co-Director, Triangle;  
Edward Strudwick, Mental Health Providers Forum

**Apologies:** Professor Gwen Bonner, University of West London;  
Dr Helen Killaspy, University College London;  
Sue Mclaughlin. Nurse Consultant, Berkshire NHS Foundation Trust

**Note-taker:** Sian Harris, Associate, Triangle

## Introduction

The Recovery Star is widely used but research on its psychometric properties and impact as an intervention has been limited. Two published studies exist. Dickens et al (2012)<sup>1</sup> report good internal consistency, low item redundancy and a two factor structure. Killaspy et al (2012)<sup>2</sup> reports good test: retest reliability, convergent validity with a social functioning measure, and high acceptability to staff and service users but below threshold inter-rater reliability. Killaspy concludes that it is not valid as a clinical outcomes measure, though the methodology of the study has been questioned.

---

<sup>1</sup> Dickens, G., Weleminsky, J., Onifade, Y., Sugarman, P. (2012) Recovery Star: validating user recovery, The Psychiatrist, 36, 45-50

<sup>2</sup> Killaspy, H., White, S. and King, M. (2012) Psychometric properties of the Mental Health Recovery Star, British Journal of Psychiatry, 201: 65- 70

In addition, the novel nature of the tool, in particular the collaborative rating approach, its dual role as a measurement tool and an intervention, and its focusing on recovery process rather than symptom severity raise substantial methodological issues in research. The purpose of this seminar was to explore the issues, agree what the focus and approach of future research on the Recovery Star should be, and plan actions required to make the research happen.

This report provides a summary of the discussion and conclusions, organised under three headings

- 1) The Recovery Star as an intervention
- 2) The Recovery Star as a measurement tool
- 3) Agreed Next Steps

In addition there are two appendices:

Appendix One: Summary of presentations made to the seminar

Appendix Two: Outline of discussion

## **1) The Recovery Star as an intervention**

Participants welcomed the fact that the Recovery Star uses a collaborative approach and it was pointed out that there is evidence from elsewhere that using an outcome measure in discussion between a worker and a service user can improve outcomes.

It was felt that the Recovery Star could help drive a paradigm shift toward a recovery orientation in mental health services and support professionals in making that shift. The primary function of the tool was seen as mobilising the agency of the service user and the worker towards achieving recovery outcomes for the service user.

It was felt that there was a need to assess its “capacity as a technology to generate improved outcomes for individuals and to improve professional practice for practitioners and teams, and better public health gains for public health care systems”

It was proposed that the Recovery Star could be seen as a ‘complex intervention’ as defined by the Medical Research Council (2008).

It was felt that a Randomised Control Trial might be an appropriate method for establishing the efficacy of the Star as a complex intervention but that before that took place it would be important that developmental work was carried out in order to:

- Outline the theoretical underpinnings of the intervention and the hypothesized mechanism of change
- Look at the degree of variation in the way it is used and implemented and how these contexts and influences might impact on outcomes. In doing this it would be important to identify what needs to be in place to implement the tool well so that positive results can be attributed to the Star rather than other features of the care environment
- Identify measures other than the Star to establish the impact of the Star and look at how the data should be analysed

- Look at service user input into the design
- Create a third edition of the Recovery Star based on feedback from users and findings of research studies

### ***Possible actions***

- Write article on theoretical underpinnings of the intervention and the hypothesized mechanism of change
- Carry out smaller studies addressing some of the questions above and draw on these to develop a third edition of the Recovery Star
- Based on this work to prepare a bid for a substantial piece of research on the effectiveness of the Recovery Star as an intervention

## **2) The Recovery Star as a measurement tool**

### **2.1) Inter-rater reliability**

The Killaspy et al (2012) study found that the Recovery Star had high test: retest reliability and convergent validity with a measure of social function but that on most of the scales the inter-rater reliability was below the 0.7 threshold. In addition Dickens et al (2012) found high internal consistency and low item redundancy. However participants noted that:

- a. A number of methodological issues had been raised about this study in the British Journal of Psychiatry including the fact that the tool had not been used collaboratively for inter-rater testing, and
- b. The results were comparatively positive and in line with those of other commonly used tools in the mental health field.

In addition some felt that it was not necessary to establish inter-rater reliability to validate the tool as it is better viewed as a “collaborative patient reported measure” rather than as a staff assessed measure. Burgess et al (2011)<sup>3</sup> have proposed 4 criteria for such tools, none of which relate to inter-rater reliability – internal consistency, concurrent validity, test-retest reliability and sensitivity to change.

Further, the view was also expressed that it would not be expected that different raters would come up with similar scores because the Star itself acts as an intervention which moves the service user along the journey of recovery. The nature and extent of that movement is dependent on the interaction between the two particular people involved and so is likely to depend on the particular service user and worker using the tool.

### ***Possible actions***

- Theory article on Recovery Star could address the question of what kind of tool the Star is and whether it is necessary or desirable to demonstrate inter-rater reliability

---

<sup>3</sup> Burgess P., Pirkis J, Coombs T, Rossen A. Assessing the value of existing recovery measures for routine use in Australian mental health services. *Aust N Z J Psychiatry* 20011; 45: 267-80

- Further work on inter-rater reliability could use written case studies in order to test the tool outside the context of an interaction, whilst ensuring that all raters have the same amount of information available to them

## **2.2) What does the tool measure?**

Some participants felt that it was not helpful to describe the Recovery Star as an 'outcomes tool'. It was suggested that the Star be described as a 'process of recovery measurement tool' rather than an outcome tool because what is measured by the tool is progress in the process of recovery. Such a measure is complementary to other types of outcome measurement.

### ***Possible actions***

- Theory article on Recovery Star could address the question of what the tool measures and propose this classification

## **2.3) How should the data be analysed?**

Peter Ford presented work carried out on the pilot data for the Spectrum Star – another version of the Outcomes Star designed for people on the autistic spectrum. This shows how a hierarchical cluster analysis can be used to determine the relative loadings of the items onto the total score. Following this, item response theory was then used to transform the ordinal data to normalised z-scores for all of the scales. This adjusts the scores so that they mean the same across the domains and could then provide a basis for further statistical testing to investigate the psychometric properties of the tool.

Such an analysis would require a large data set of at least 500. It was agreed that it would be useful to apply this approach to the Recovery Star and that substantial data sets already exist to enable this analysis

### ***Possible actions***

- Draw on Recovery Star data already held on the Star Online and the Mental Health Providers Forum database
- Investigate availability of further datasets eg CQUIN data

## **3) Agreed Next Steps**

### ***Smaller scale studies***

Undertake a number of smaller studies to address issues explored in the seminar. These could be carried out by Masters Psychology students and as part of existing research programmes

Action: Sue Holltum, Jerry Burgess and Richard Byng

### ***Third edition of the Recovery Star***

Alongside the smaller studies, develop a third edition of the Recovery Star

Action: Joy MacKeith

### ***Larger study***

Once the small studies have been carried out and published, prepare a funding bid for a large scale grant to research the Recovery Star as an intervention

Action: Sue Holltum, Jerry Burgess and Richard Byng (Chris Evans also agreed to provide support in relation to sources of funding)

### ***Special edition of a Journal***

Approach relevant journals to suggest a special edition focusing theory and practice pieces on the Recovery Star.

Action: Mark Freestone

### ***Analysis of existing Recovery Star data***

Draw together existing data on Triangle's Star Online, the Mental Health Providers Forum database and CQUIN data to provide the substantial data set needed to convert the data into parametric data and carry out further analysis on this data

Action: Peter Ford and Jerry Burgess (with possible input from Nottingham Evaluation Centre – Chris Evans to explore)

### ***On-going coordination***

A report of the seminar will be produced and there will be on-going liaison between members of the group to take forward the proposals outlined here

Action: Joy MacKeith

## Appendix One: Summary of Presentations

### Background of the Recovery Star and Issues for Discussion *Joy MacKeith*

---

- ❖ Joy discussed the research already available on the Recovery Star as a key work tool:
  - Anecdotal evidence suggests that both service users and workers find it useful, with workers suggesting that it helps new conversations to happen that wouldn't have transpired otherwise.
  - Killaspy et al (2012) found that 85% of service users and 85% of workers found the Star useful or very useful.
  - Furthermore, at an organisational level, anecdotal evidence suggests that the Star is helpful for translating recovery values into practise, and for communicating across multi-disciplinary teams.
  
- ❖ There is also research on the Star as an outcomes tool:
  - The Mental Health Providers Forum and St Andrews Academic Centre found that 7 of the 10 scales were normally distributed, the internal consistency was high, item redundancy was low, and responsiveness was good, though refinement was needed on some scales.
  - Killaspy et al (2012) found that test-retest reliability was high for collaborative use of the star, though inter-rater reliability was low. They also found convergent validity with a measure of social functioning but not a measure of recovery.
  - The Killaspy et al (2012) results on inter-rater reliability are below the normally accepted threshold. However, the research is flawed methodologically; the paper only tests independent service worker ratings, and not collaborative ratings as the Star is intended to be used.
  
- ❖ Based on the current research available, two of areas of research are proposed going forwards:
  - Testing collaborative inter-rater reliability and convergent validity.
  - Research into the Star's use as a key work intervention, including its impact on service user, workers and organisational.
  
- ❖ Joy also identified three key issues for discussion during the seminar:
  1. ***Can a key-work tool also be an outcomes tool?***

It is suggested that the two functions are supportive of each other in four ways:

- The specificity required in outcomes measures provides concrete values and behaviours for service workers and users to visualise and target.
- The step-change breakdown used in outcomes measures makes change in the service user more manageable.
- Service users like using it, so no need to collect separate outcomes data.
- It aligns the needs of commissioners and service workers.

However, it is acknowledged that the two functions are also counteractive as the specificity needed in outcomes measures would require a lot of detail resulting in a long tool that service users may be reluctant to complete and engage with.

## **2. *Is it meaningful to focus on relationship with the problem rather than problem severity?***

It is suggested that front-line workers identified relationship with the problem as more important, as it is what determines intervention and likely progress.

## **3. *Is a collaborative approach to measurement meaningful and valid?***

It is suggested that this is an advantageous approach as it feeds both service user and professional perspectives into the picture providing a fuller and more rounded view than one of these perspectives on its own. However, it is acknowledged that combining the two perspectives in one measure makes it difficult to identify the interaction between the two.

### **Presentation: Finding a Parametric Solution *Peter Ford.***

---

Peter pointed out that the Star produces non-parametric ordinal data, which is not normally distributed. This produces inaccurate conclusions when using parametric statistics on the data, such as when used by Killaspy et al (2012).

The solution may be to use hierarchical cluster analysis to determine the relative loadings of the items onto the total score. Following this, item response theory would be used to transform the ordinal data to normalised z-scores for all of the scales. This would adjust the scores so that they mean the same across the domains. Such an analysis would require a large data set of at least 500.

### **Presentation: New Research Plan *Sue Holttum.***

---

Sue presented four possible studies that would aid development of the Recovery Star:

- ❖ Interviews with staff and service users whilst using the Recovery Star to make ratings. Attention will be to the kind of factors that are taken into account when producing a score, and how these factors differ between service workers and service users, and additionally how they differ between independent and collaborative ratings. This information may help within with defining version 3 of the Star.
- ❖ Testing inter-rater reliability in a situation which removes the inevitable disparities that arise when two different service workers collaborate with one service user. This will be done by getting two independent raters to listen to a collaborative rating, and decide the outcome score based on the collaboration.
- ❖ A new factor analysis utilising the large archive of data available.
- ❖ Pilot combined validation study and intervention trial of version 3 of the Star. It is expected that use of the Star in key work elicits movement towards positive change.

**Presentation: Evaluation of the Recovery Star as an Outcome Generator** *Richard Byng*

---

Richard suggests that the Recovery Star is a complex intervention, and as such there are three issues to cover when producing research for its evaluation:

- ❖ Developing the method – a randomised design would be preferable, and outcomes such as social inclusion, mental health and coping strategies should be used. The control group would need a case management tool to replace the function of the Recovery Star in the experimental group.
- ❖ Feasibility – suitable client groups, teams and services will need to be identified. Individuals willing to take part in the research would also need to be found.
- ❖ Developing the intervention – work needs to be done to ensure the core elements of the Recovery Star are reproduced consistently by practitioners so that the positive result of the intervention can be attributable to the Recovery Star rather than other properties of their care.



## Appendix Two: Outline of discussion

The key points discussed were as follows:

---

- ❖ Clarifying the use of parametric vs non-parametric data:
  - By converting the data into parametric data it is more likely to produce reliable results when tested.
  - It is important that the Star is used the same way by the service providers, as its visual 'star' form is helpful in key work. Additionally, it would be hindering to ask key workers to convert the data just for key work purposes. Therefore, conversion should only be used when testing the reliability of the tool as an outcome measure.
  - It is established that the data that is available and has already been used in research, such as the data used by Killaspy et al (2012), could be converted into z-scores and re-tested for reliability.
  
- ❖ Usefulness of a control group:
  - Including a control group of a normative population may help get a more rounded picture of how scoring on the Star sits among the normal population. This may be information of interest to commissioners. A control group could also be helpful to show service users where they could be, and that other people experience these problems, though are coping well outside the service.
  - However, the Star was developed with the idea in mind that there would be further development beyond the top score of 10. Hence it would be expected that in a population of people not using mental health services most people would score 10 on most scales and the Star would not be very discriminating in this population
  
- ❖ Amending the 'Work' scale:
  - The scale does not allow for people who are in work, but it is the work that is causing their illness.
  - The work scale could be modified in the same way that it is used in the Drug and Alcohol Star, which gives an option for being in work and struggling.
  
- ❖ Producing version 3 of the Recovery Star:
  - There is currently no time frame for completing the new version of the Recovery Star, or the associated research.

- Preferably the new version would be completed during the early stages of research, so that validity and reliability testing can be completed on the new version rather than the older version.
  - Producing the new version within the context of a piece of research may present funding issues, as it may be assumed that there is a lack of confidence in the tool as an intervention.
- ❖ Using collaborative scoring:
- There are some anecdotal issues with using a collaborative score, such as how do you deal with a situation where there is different perceptions between the worker and the service user.
  - This is especially important if a worker knows they have to produce one score in order for the tool to be used as an outcome measure.
  - Positively, collaborating can improve the outcome, as an intervention it works with the client to help them to understand where they are in reality.
  - In training for use of the Recovery Star it is made clear that the relationship between the worker and the user is paramount. Open disagreement about the score should be avoided if it will affect this relationship.
- ❖ Key work tool vs. outcome measure, and inter-rater performance:
- The research so far suggests that the Recovery Star does not perform badly as a key work tool or as an outcome measure. More research would be needed to decide whether to continue pursuing both of these elements.
  - Poor inter-rater reliability as reported by Killaspy et al (2012) should not be an issue, as the nature of a collaborative measure means that with each different worker the score is likely to be different based on their interaction with the service user.
  - Additionally, Killaspy et al (2012) used two ratings which were up to a month apart, in which time the ratings are likely to be different due to real changes in the service user.
- ❖ Reconsidering the name of the tool:
- Currently the Recovery Star is referred to as an outcome tool.
  - Changing this label may alter how it is viewed and ensure that it is used by service providers in the way that it is intended.
- ❖ The Recovery Star as an intervention:

- When testing the Star as an intervention the control group would have to be receiving normal case management, whilst the experimental group would use the Star within their case management.
  - The success of the intervention could be measured against mental health symptoms or process of recovery.
- ❖ Enhancing normal case work:
- It may be helpful to know what happens in normal case management, as it may be that using the Star aids more productive case management.
  - There were concerns that if the Star were imposed on workers in order to aid case management this could lead to issues where case workers are reluctant to use the Star as they may feel it is cumbersome and restrictive.