

UNIVERSITY OF WALES INSTITUTE, CARDIFF



ATHROFA PRIFYSGOL CYMRU, CAERDYDD

Research Report

Validating the Outcomes Star as a data collection tool

October 2009

Catherine Boswell and Deiniol Skillicorn

University of Wales Institute Cardiff

Table of Contents

1. Executive summary.....	3
1.1 What was the project?	3
1.2 Methodology.....	3
1.3 Why is consistency important?	3
1.4 How consistently is the tool used?	4
1.5 What factors affect its consistency?	4
1.6 Do the scores correlate to hard outcomes?.....	4
1.7 What can organisations do to be more consistent?.....	4
1.8 What can commissioners learn from this?.....	4
1.9 What are the next steps?.....	5
2. Introduction	6
2.1 Research methodology.....	6
2.1.1 Consistency	6
2.1.2 Correlation between ‘hard’ and ‘soft’ outcomes	7
3. The findings	8
3.1 Consistency of use of Outcomes Star	8
3.1.1 What the variation is	8
3.1.2 Factors influencing variation	10
Conservatism	11
Using the ladders in the guide.....	12
Training	12
3.1.3 Qualitative Issues	13
Issue: clients unable to progress.....	13
Issue: key-work or data collection tool?	14
Issue: non-engaging clients.....	14
Issue: wording of ladders	14
Issue: differences of opinion.....	15
3.1.4 Ways identified to overcome these issues.....	15
Formal training	15
The need for different Stars.....	16
Recognition of different uses.....	16
3.2 Correlation between ‘hard’ and ‘soft’ outcomes.....	16
4. Learning and recommendations.....	19
4.1 Consistency.....	19
4.1.1 For organisations.....	19
4.1.2 For commissioners	19
4.1.3 For the Star.....	20
4.2 Correlation to hard outcomes	20

1. Executive summary

1.1 What was the project?

Homeless Link is the national membership organisation for frontline homelessness agencies in England. In April 2009 we commissioned University of Wales Institute, Cardiff to validate, through research, the effectiveness of the Outcomes Star¹ as a data collection tool. This was to determine:

- a) whether the tool is used consistently across different workers, projects and organisations
- b) whether progress in 'soft' outcomes, as measured by the Star, correlates to the achievement of 'hard' outcomes.

This work was funded by the Big Lottery.

1.2 Methodology

Consistency:

Organisations that regularly use the Outcomes Star were asked to take part in an on-line survey. This comprised of two case-study scenarios, based upon which front-line staff completed an Outcomes Star Chart. They also answered a small number of questions on factors such as training they had received and the length of time they had been using the Star. This survey was then followed up with focus groups to discuss any variance.

Correlation to hard outcomes:

The researchers looked for tangible positive outcomes for service users that were evidenced by both soft and hard data. A number of organisations that have been using the Outcomes Star since April 2007 as part of the Link client recording system were asked to participate in the research, and data was extracted and analysed for all those service users who have moved on and for whom at least two Outcomes Stars have been completed. This data was analysed using the statistical analysis package SPSS, to establish whether there was a statistically reliable relationship between the data as measured by the two tools.

1.3 Why is consistency important?

Consistency of scoring on the Star is important because without it no meaningful analysis can take place of the data from different workers, projects and organisations on anything other than an individual basis. When scoring is recorded consistently, the data can be examined and compared to form the basis of service evaluation and improvement, as well as to provide meaningful evidence of the difference that services are making in people's lives.

¹ The Outcomes Star is a comprehensive tool for measuring the outcomes of work with homeless people. It was developed by Triangle Consulting, originally for the homelessness charity, St Mungo's, and has subsequently been widely tested and revised with funding from the London Housing Foundation. Homeless Link is now managing the development and dissemination of the Outcomes Star tools and resources and last year republished the Outcomes Star guidance manuals.

1.4 How consistently is the tool used?

The research showed that the variance in scoring between workers within a project is currently between 1 and 3 points on the Star. This can be reduced further still by the factors discussed below. However, between organisations the variance can be as much as 3 to 6 points on the Star.

1.5 What factors affect its consistency?

There appears to be a number of key factors which affect the consistency of use of the tool. These are:

- Formal training – this may help to improve consistency and is likely to make staff more confident to use their scoring (rather than compromise or use the clients’).
- Length of time using the Star – as workers become more familiar with the tool through use, they begin to record scores more consistently.
- More conservative scores appear to tend to be more consistent, and formal training might result in more conservative scores.
- Using the ladders as provided in the guidance results in more consistent scoring.
- The variance in scoring is greater on some scales of the Star than on others – this could be down to interpretation of wording on particular scales.

1.6 Do the scores correlate to hard outcomes?

The analysis produced a result that is counter-intuitive, and consequently, it might be useful to revisit the Link data when there has been an opportunity to disseminate good practice in completing the Outcomes Star. However, there is currently no evidence that there is a statistical linkage between the progress recorded by the Outcomes Star (the ‘soft’ outcomes) and that recorded by the Link database (the ‘hard’ outcomes).

1.7 What can organisations do to be more consistent?

There are several steps that can be taken by organisations to ensure that the data you collect is more consistent. These are as follows:

- Discuss the Star at team meetings to develop a shared understanding across teams.
- Enable formal training to take place wherever possible.
- Allow time for implementation of the tool and discuss its use frequently during early months.
- Encourage staff to use the ladders as part of their key-working to help with scoring, or regularly revisit them to refresh their knowledge.

1.8 What can commissioners learn from this?

Commissioners need to be aware that there can be a variation in scores between organisations of between 3 and 6 points. You should bear this in mind if using Outcomes Star data to compare different services.

You should also be aware that the tool has two different purposes – one is to improve key-working² and the other is to improve data collection. You should therefore be very clear on why you want organisations to use the tool, as focusing on one of these aims will have implications for the other.

You can also support the organisations you commission to record data more consistently using the Outcomes Star by sharing the good practice recommendations above.

1.9 What are the next steps?

Homeless Link will disseminate these findings to those using the Outcomes Star and to organisations that could benefit from the tool. We will also disseminate good practice in using the tool consistently to ensure an improvement in its use³.

Homeless Link will continue to work with Triangle Consulting to review the ladders where interpretation of wording may be an issue to make this clearer, as well as to link organisations to other specialist Stars where appropriate.

Finally, Homeless Link will develop proposals to undertake more detailed research on the correlation of the soft outcomes as recorded by the Star and harder outcomes.

² See Impact and Good Practice Research Report at <http://www.homelessoutcomes.org.uk/starimpact.aspx>

³ To see good practice, refer to the User Guide at http://www.homelessoutcomes.org.uk/resources/1/Outcomes%20Manuals%202nd%20Ed/Star_organisation_guide_2ndEd.pdf

2. Introduction

The Outcomes Star is a comprehensive tool for measuring the outcomes of work with homeless people. It was developed by Triangle Consulting, originally for the homelessness charity, St Mungo's, and has subsequently been widely tested and revised with funding from the London Housing Foundation. The tool is currently in use by a number of homelessness organisations in the UK. Increasingly there is awareness on the part of statutory agencies and funders that measuring progress for individuals is complex and requires a greater degree of sophistication than simple numerical data. The Star is a tool that can be used to evidence the 'soft' outcomes for clients, but there is some (albeit limited) evidence that there is a correlation between positive progress as measured by the Star and 'hard' outcomes⁴. The purpose of this validation research is twofold

1. To establish whether the Outcomes Star is being used in a consistent manner by within and between the agencies
2. To see whether there is a clear correlation between the progression made by individuals as measured by the soft outcomes from the Star, and positive hard outcomes, primarily considering successful tangible progression.

2.1 Research methodology

The choice of methodology has been led by the information and tools available for analysis, and also by the earlier work undertaken by Triangle Consulting; primarily the research report produced in September 2008⁵. There are two existing sources of information that can be drawn upon. One is the Outcomes Star itself, which provides valuable data, the other is the Outcomes monitoring data held by Homeless Link on behalf of a range of member organisations. These sources of data have been supplemented by group discussions with front-line staff.

To ensure that the available evidence is fully examined, a multi-dimensional approach has been adopted to tackle each of the two aims of the research.

2.1.1 Consistency

In order to compare 'like with like', two case-study scenarios were devised, and workers were asked to complete an Outcomes Star Chart based on these scenarios. Organisations were asked to participate via an on-line survey. The survey was aimed at front-line staff that regularly use the Outcomes Star, and a total of 33 fully completed surveys were submitted. The two case-studies ensured that all scales of the Star were covered between them. Participants were also asked for the following simple additional information:

- How long have you been using the Star?
- Which client group do you predominantly work with?

⁴ Burns, S. MacKeith, J. and Graham, K. (Sept 2008) Using the Outcomes Star: Impact and good practice. Triangle Consulting research report for Homeless Link p 17

⁵ Ibid

- Who provided any training you received about how to use the Outcomes Star?
- Did you use the ladders provided in the guidance documents when giving the scores on the case studies?
- Approximately how frequently do you complete an Outcomes Star with each client?

However, in order to achieve a maximum response the survey was kept short and succinct.

This was then followed up with focus group discussions involving front-line staff from two of the participating organisations. The meetings were intended to explore in detail how staff make use of the Outcomes Star, the process, and to examine variances in the case-studies completed as part of the survey. The focus group attendees were drawn from two contrasting organisations in different parts of the country, both being from the voluntary sector, one a single project that is part of a national network, and the other a large organisation with geographically dispersed multiple projects.

2.1.2 Correlation between 'hard' and 'soft' outcomes

The question posed is whether there is any evidence that soft outcomes as measured by the Outcomes Star have any direct correlation to hard outcomes, primarily being measured as positive and tangible results for the client directly related to the support service provided by the organisation. In this instance, the positive outcome is identified from the data held by Homeless Link. The indicators have been mapped as closely as possible against the Outcomes Star.

The approach adopted is that of Impact Evaluation, which in this context means we are looking for a tangible positive outcome for the service user, which is evidenced by both hard and soft data.

In order to track the progress of the individual service user, the actual or tangible outcomes taken from the LINK data, were mapped against the effects of the support provided as measured by the Outcomes Star.

Housing outcomes were analysed, looking at:

- Was the move-on planned or unplanned?
- Did the move result in greater independence?

A number of organisations that have been using the Outcomes Star since April 2007 were asked to participate in the research, and data was extracted and analysed for all those service users who have moved on and for whom at least two Outcomes Stars have been completed.

The data was analysed using the statistical analysis package SPSS, to establish whether there is a statistically reliable relationship between the data as measured by the two tools.

3. The findings

3.1 Consistency of use of Outcomes Star

The respondents

- Thirty three respondents completed the on-line survey, representing fifteen different organisations, geographically dispersed throughout the UK with most regions being represented.
- Three organisations had multiple respondents ranging from four to seven (these have been identified in this report as Projects A, B, and C) while the remainder of organisations had one or two staff completing the survey.
- Two of these three multiple respondent organisations were also involved in the focus group discussions.
- Most people had been using the Star for at least six months
 - The exception to this was Project C, where none had been using it for more than three months.

Caveat

The number of responses received through the online survey was too low to give results that have statistical validity. This should be borne in mind when reading this report and the information within it should not be used out of context or to draw definite conclusions. However, the collected data does help us begin to identify trends and indicates issues that could benefit from further exploration.

3.1.1 What the variation is (**between and within organisations**)

Figure 1 below shows the variation (range) of responses for each scale of the Star across the twenty four respondents who used the ladder⁶ to complete the Star. The amount of variation can range from 0 (i.e. no variation) to 10 (i.e. maximum variation). Smaller variability might indicate more consistent responding. The figure below shows that the three scales with the most variation in responses are 'self care and living skills', 'drug and alcohol misuse', and 'offending'. This might indicate that responses in these scales are more affected by subjective opinion, or that the ladders aren't as clear.

⁶ See methodology section 2.1.1

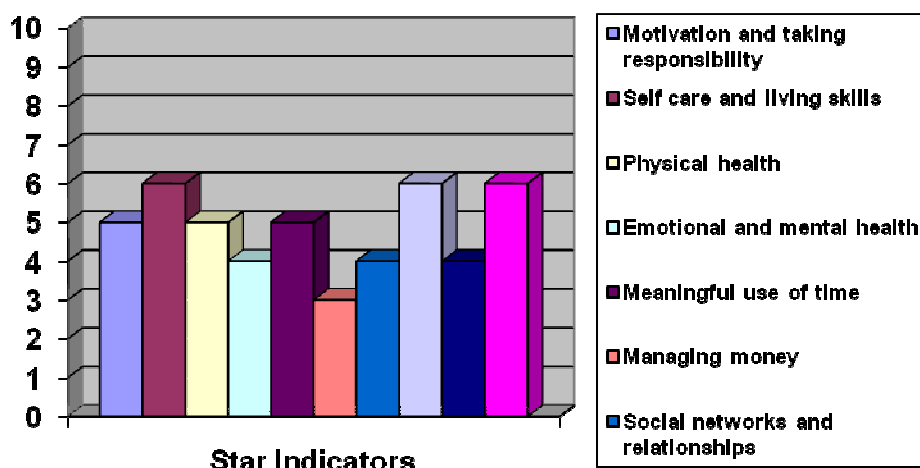


Figure 1. Range of responses for the twenty four participants who used the ladder to complete the Star

The three organisations with multiple respondents were looked at in more detail to see how much within project variation in responses there was between staff. Table 1 below shows the variation (range) in responses of the three multiple respondent organisations for each of the ten scales of the Star.

Table 1. Range in responses of the three organisations with multiple respondents for each of the ten scales of the star

Star scale	Project A	Project B	Project C
Motivation and taking responsibility	1	3	2
Self care and living skills	4	2	4
Physical health	4	3	2
Emotional and mental health	3	2	3
Meaningful use of time	4	3	1
Managing money	2	2	3
Social networks and relationships	2	0	4
Drug and alcohol misuse	4	1	6
Managing tenancy and accommodation	1	0	4
Offending	3	1	2

For each of the ten scales of the star the largest amount of variation is highlighted to see if there is a pattern. The scale is not highlighted where there is no difference between the three organisations, or where 2 of the 3 organisations have the same score. A variation in response between 1 and 3 points was common and some variation is to be expected. However the 4 to

6 point variation might suggest a more fundamental difference in interpretation. The organisation with the lowest level of variation was Project B. Beyond this, however, there does not appear to be any obvious pattern within the data.

This data was further examined using the means and standard deviation to give a more accurate reflection of variance. Figure 2 below presents the mean (or average value) with the error bars representing the standard deviation (or the amount of variability). As an example, the mean response for 'drug and alcohol misuse' from Project C is 3.5 with a standard deviation of 2.5. For the same variable, Project A had a mean response of 3.9, and a smaller standard deviation of 1.2. In other words, the larger standard deviation indicates that the staff in Project C were less consistent in their responses than the staff in Project A.

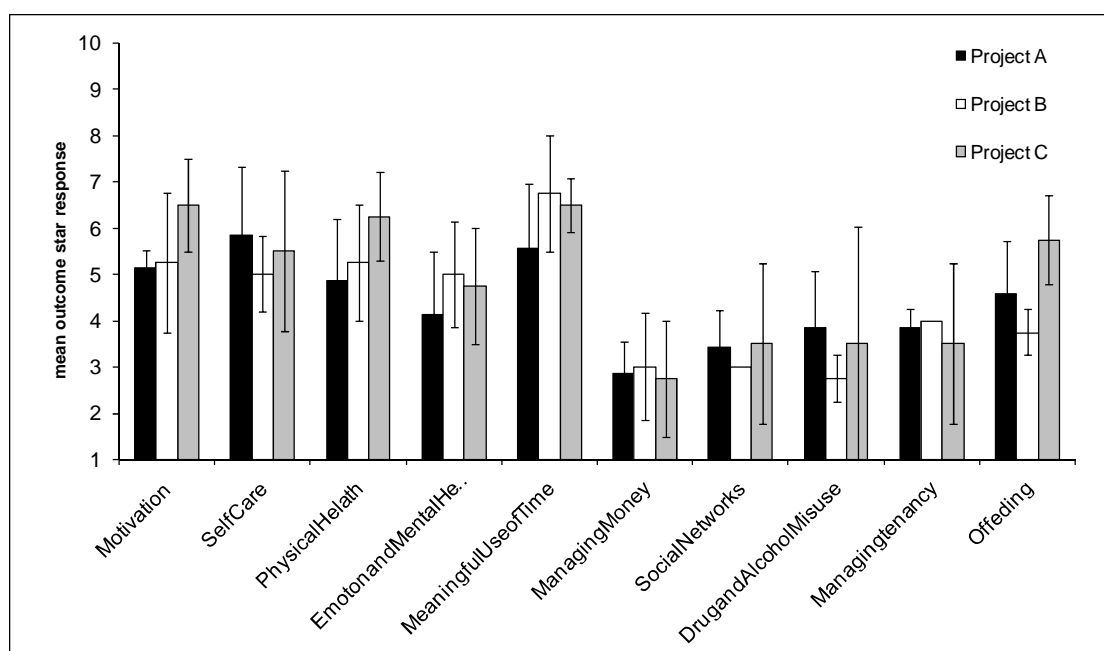


Figure 2. Participants' responses for the 3 top responding organisations (Figure includes the standard deviation)

3.1.2 Factors influencing variation

The factors that may influence this are as follows:

- The four staff in Project C have been using the Outcomes Star for less than 3 months and have received a mixture of formal and informal training in the use of the Star
- The four staff in project B had been using the Outcomes Star for a range of periods (i.e. greater than 12 months, between 6 and 12 months, and less than 3 months) and had all received informal training in the use of the Outcomes Star from colleagues

- The seven staff in Project A have been using the Outcomes Star for a minimum of 6 months and have all received formal external training in its use.

Project A is therefore the most representative of the effects of formal training in the use of the outcome stars. For this reason, the following tables will specifically examine the means and standard deviations of Project A. The data for Project B and C will be included for comparison but will not be commented on due to the inconsistency in type of training received, the varying amount of time using the Outcome Stars and crucially their small size.

Table 2 below indicates that, compared to project B and C, the smallest standard deviation in Project A is for 'motivation' and 'managing money, which might reflect a greater level of consistency in response for these two scales.

Table 2. Standard deviation of responses of project A, B and C to each scale of the outcome star.

	Project A (n=7)	Project B (n=4)	Project C (n=4)
Motivation	0.4	1.5	1.0
Self Care	1.5	0.8	1.7
Physical Health	1.3	1.3	1.0
Emotion and Mental Health	1.3	1.2	1.3
Meaningful Use of Time	1.4	1.3	0.6
Managing Money	0.7	1.2	1.3
Social Networks	0.8	0.0	1.7
Drug and Alcohol Misuse	1.2	0.5	2.5
Managing tenancy	0.4	0.0	1.7
Offending	1.1	0.5	1.0

Conservatism

Compared to Project B and C, the lowest mean score for Project A were on the three scales of 'motivation' 'physical health' and 'emotional and mental health' (see table below). This could be indicating that Project A are responded with a greater level of conservatism in their assessment of where a service user currently stands on those outcome stars scales.

Table 3. Mean responses of project A, B and C to each scale of the outcome star.

	Project A (n=7)	Project B (n=4)	Project C (n=4)
Motivation	5.1	5.3	6.5
Self Care	5.9	5.0	5.5
Physical Health	4.9	5.3	6.3
Emotion and Mental Health	4.1	5.0	4.8
Meaningful Use of Time	5.6	6.8	6.5
Managing Money	2.9	3.0	2.8
Social Networks	3.4	3.0	3.5
Drug and Alcohol Misuse	3.9	2.8	3.5
Managing tenancy	3.9	4.0	3.5
Offending	4.6	3.8	5.8

Using the ladders in the guide

The next table displays the means and standard deviation for each scale of the Star scored by those who made use of the ladders and those who did not use or only partly used the ladders. This data relates only to those respondents who made use of the ladders, it does not reflect whether they had received formal training in how to make effective use of the Outcomes Star.

Table 4. Mean and standard deviation of participants' responses for those who received formal training and those who received informal or no training

	Yes (n=24)		No/Partly (n=6)	
	Mean	SD	Mean	SD
Motivation	5.42	1.06	6.17	1.72
Self Care	5.21	1.41	5.00	1.10
Physical Health	5.17	1.49	5.50	1.05
Emotional and Mental Health	4.46	1.32	5.17	1.33
Meaningful Use of Time	5.87	1.36	6.00	1.41
Managing Money	2.71	0.81	2.67	0.82
Social Networks	3.21	0.88	3.00	1.10
Drug and Alcohol Misuse	3.50	1.44	3.40	1.14
Managing tenancy	3.92	0.93	3.17	1.17
Offending	4.58	1.35	3.50	1.64

An examination of responses of those who used the ladders with those who either did not or just partly used the ladders indicates that there is a marginally smaller level of response variation for seven of the Outcomes Star scales amongst those who used the ladders to complete the Star (see table 4 above). These observations must again be treated with caution, as only six respondents did not use, or only partly used the ladders compared to twenty four respondents who did use the ladders.

Training

Table 5. Mean and standard deviation of participants' responses for those who received formal training and those who received informal or no training

	Formal (n=15)		Informal/none (n=15)	
	Mean	SD	Mean	SD
Motivation	5.4	1.0	5.7	1.4
Self Care	5.2	1.4	5.1	1.4
Physical Health	4.9	1.4	5.6	1.4
Emotion and Mental Health	4.3	1.5	4.9	1.1
Meaningful Use of Time	5.7	1.4	6.1	1.3
Managing Money	2.7	0.7	2.7	0.9
Social Networks	3.1	0.9	3.2	0.9
Drug and Alcohol Misuse	3.5	1.2	3.5	1.6
Managing tenancy	3.6	1.1	3.9	1.0
Offending	3.9	1.6	4.8	1.1

Looking at all thirty three respondents, table 5 above presents the mean and standard deviation in responses for each scale of the Star comparing staff who received formal training with those who either had informal - such as from a colleague - or no training. While there does not appear to be any pattern of difference between the two groups in terms of consistency, there is some evidence of a tendency amongst those who have received formal training to use a lower score in many scales of the Star as compared with those participants who had received informal or no training.

3.1.3 Qualitative Issues

The focus of this section will be on the two multiple response organisations, whose staff attended a group discussion session to explore some of the issues in more detail.

Issue: clients unable to progress

Looking at 'self care and living skills', Project B, who work with a range of clients, including some of the most chaotic people who have spent long periods of time street homeless, commented that the ladder is difficult to apply to some of their residents. Participants noted how some people may not move past level 2, yet their achievements, for them are significant. One resident, for example, had been with them for a number of years, and although the Star reflected a lack of progress, this was not a helpful representation of reality. This was epitomised by the comments:

For some, being able to make a meal once a week, that isn't from a chip shop and isn't a pot noodle, it's a ham sandwich, is a massive deal, but the Star doesn't pick up the significance of a ham sandwich

One guy has been here for 6 years and is still only getting 2s, but it still has to be done (the Star) every six months. He's not improving and has complex needs. Washing his clothes is a major step

Somebody having moved only one step on a statistics sheet looks really poor, but if someone comes downstairs every morning having brushed their teeth without you having to tell the, even though they've put dirty clothes back on, that is amazing

What is highlighted here, is the tension between trying to reflect progress - which is what the Star is designed to do - in the context of some residents who may well have progressed from the point they were at when they first came to the project, but who may not be able or willing to move any further. This is overlain by the need to demonstrate that a project is effective through the statistical information produced.

Key point – the Star is difficult to use with clients who have made progress but are unwilling/unable to move further

Issue: key-work or data collection tool?

However, the overwhelming majority of people found the Star to be a valuable tool to use with clients, though there are differences in emphases:

What the workers get out of it, what the service user gets out of it, and what management get out of it are all very different things

Key point – a balance is needed between it being a key-working tool and a data collection tool

Issue: non-engaging clients

In discussion, it became apparent that a small number of service users have little or no engagement with the Star and although it plays a useful role in providing a structure for the key-worker and helps to open up areas for discussion, they do not have any involvement with it, hence the marking on the ladder reflects the view of the key-worker alone.

To be honest, sometimes it is an academic exercise, and it's purely from the perspective of the key-worker

Key point – the Star can be difficult to use with clients who have little or no engagement with the process

Issue: wording of ladders

The scale of drug and alcohol misuse caused most difficulty when working with young people, where staff found the wording of the stages of the ladder sometimes difficult to apply to their service users.

It's geared towards whether you're following a programme, most of our young people probably smoke cannabis, but they're not following a programme so it can be difficult to help them see where they are (on the scale). They say I can't be a 7 because I haven't got a drug worker, and it's difficult for them to see how it fits with them.

Main problem is the drug and alcohol wording.

Wording seems more specific to Class A long-term usage

Wording is quite contradictory; sometimes you'll see the same statement in two places which throws you

For young people, the offending category was thought to overlap with drug and alcohol misuse as many young people made habitual usage of cannabis.

If they're smoking it's against the law and also in breach of their tenancy agreement if they smoke in the hostel, but they still don't think they need support

The reasons for discrepancies are similar, inasmuch as the service user would not feel that there was an issue and consequently saw an appropriate

point on the ladder as being a 10, whereas the support worker would look to scoring much lower at a 1 or 2. The eventual score would vary, depending upon whether the position adopted was one of compromise (5 or 6 on the ladder), reflected the support worker's view, or that of the service user. It should be noted that the position adopted was influenced by the type of training received, which will be dealt with later in the report⁷.

Another scale that some staff found difficult to apply to young people, is emotional and mental health, as although many of their clients had emotional problems, they were unlikely to be linked into formal mental health services, and struggled to relate some of the wording to their own situation.

Key point - the wording of the Star is not always appropriate to some clients (e.g. young people, or those who are very chaotic)

Issue: differences of opinion

Those project workers who had received formal external training were seemingly more confident in dealing with discrepancies and were less likely to enter a compromise score on the Star. The approach taken, as recommended in the training received, was to use the Star to reflect the viewpoint of the staff member, and to reflect the views of the service user in the support plan. This, it should be stressed, would be the outcome where no reasonable point of compromise could be found, for example:

I had a situation on alcohol and drug misuse where they thought they were a 10, and when we delved into it, I thought they were more like a 1. They were indulging in massive binge drinking, downing neat bottles of vodka and passing out in the street, but that wasn't an issue for that young person. I said I'd mark him down as a 1, but in the support plan explain why he thought he was a 10

In contrast, other people were more likely to reflect the view of the service user in the Star

If I do it (the Star) with a client, the eventual score will be theirs rather than mine

Key point – there can be some uncertainty over whose score to use, should it be a compromise, the client's perception or that of the worker?

3.1.4 Ways identified to overcome these issues

Formal training

There was a general consensus that there would always be some degree of variation, but also an acknowledgement that training was important to

⁷ See Issue: differences of opinion

minimise discrepancies and maximise the validity of the Star data for statistical purposes. One organisation was now putting formal training in place, the need having been highlighted by some of the scoring that had been occurring on the Star:

There were lots of 10s, and we were saying that people shouldn't be in these projects if they were 10s

Formal training is likely to make people more confident to use their scoring (rather than compromise or use the clients)

The need for different Stars

Although all the organisations were making use of the generic Star designed for homelessness projects, there was a view that a more tailored model for young people would help to improve consistency. The scales already identified above where there are some difficulties are drug and alcohol use, and emotional and mental health. Many young people had problems in these scales but the nature of these was thought to be of a different quality from the position of older people, and not as clearly reflected in the wording of the Star. The most chaotic of clients was the other group where difficulties were raised, partly because their ability to connect with the Star (or any other tool) was very limited, but also because progress was often so restricted that people might be on the same low point for a very long time. The solution here may be to consider a Star which is of more relevance to those people whose progress may need to be measured differently. The other groups where there may be similar issues are likely to be people with learning difficulties and frail older people.

It might be worthwhile to consider having different Stars tailored to meet needs of different client groups.

Recognition of different uses

There is also a need to recognise that it is a useful tool with clients – but there should be an understanding that different people want to use it for different things.

3.2 Correlation between 'hard' and 'soft' outcomes

The 609 Link data cases were first examined to identify those cases with a successful 'hard' outcome and those with an unsuccessful 'hard' outcome. These were defined as those who left the service under planned circumstances (i.e. successful hard outcome) and those who left the service under unplanned circumstances. The examination identified 87 'planned'

move on cases and 15 ‘unplanned’ move on cases (see table 6). These groups are examined in further detail below.

Table 6. Number of planned and unplanned cases

	Frequency
Planned	87
Unplanned	15
Total	102

In order to track changes in Outcome Star scores, a ‘difference score’ was calculated between the first and last Outcome Star assessment for each individual case. A positive score would indicate improvement, a score of 0 would indicate no change and a negative score would indicate deterioration. So, if the Outcome Star is predictive of good outcomes, or what was defined as a good (i.e. a planned move on) then this group should have positive score for some or all Outcome Star variables. Similarly, if it is predictive of bad outcomes (i.e. an unplanned move on) then we should see a negative score or no change.

Table 7 below presents the percentage of valid and missing data for each of the Outcome Star scales for both the planned and unplanned groups. Missing data is the result of absent Link data at either the first or last Outcome Star assessment thereby preventing the calculation of a difference Outcome Star change score.

Table 7. Percentage of valid and missing cases for each of the Outcome Star scales for the planned and unplanned groups

	Planned (n=87)		Unplanned (n=15)		Total	
	Valid	Missing	Valid	Missing	Valid	Missing
Motivation	72	28	100	0	76	24
Living Skills	77	23	80	20	77	23
Social	61	39	80	20	64	36
Substance	67	33	93	7	71	29
Physical	67	33	80	20	69	31
Emotional	70	30	87	13	73	27
Use of time	69	31	80	20	71	29
Managing	72	28	100	0	76	24
Offending	77	23	80	20	77	23

The table indicates that the percentage of missing data for the whole sample across the ten Outcome Star scales range from 23 to 36 percent. It further indicates that missing data from the planned group ranges from 23 to 39 percent while those for the unplanned group range from 0 to 20 percent. It is important to note the difference in sample size between the planned and

unplanned groups while considering the difference in the proportion of missing data.

Table 8 below, presents the means and standard deviation of the Outcome Star 'difference' change score of each of the Outcome Star scales for both the planned and unplanned groups.

Table 8. Means and standard deviations of the Outcome Star 'change score' for each of the Outcome Star scales for both the planned and unplanned groups

	planned		unplanned	
	mean	sd	mean	sd
Motivation	0.1	1.3	0.3	1.0
Living Skills	0.3	1.5	0.1	2.7
Social	0.4	1.5	0.7	1.2
Substance	0.1	1.8	0.4	1.3
Physical	0.3	1.4	0.3	1.8
Emotional	0.2	1.2	0.7	1.2
Use of time	0.1	1.4	0.3	1.0
Managing	0.3	1.7	1.1	1.4
Offending	0.1	1.3	0.3	1.4

(sd = standard deviation)

It indicates that, on average, none of the Outcome Star scores changed by more than one point on the scale between initial and final assessment. This was the case for both the planned and unplanned groups with one exception. The unplanned group produced a positive change of one Outcome Star scale point for the variable Managing. However, this is counter predication as it might be expected that the group with an unsuccessful hard outcome (i.e. an unplanned move on) might present with deterioration or no change in function as measured by the Outcome Star. Most standard deviations do not exceed 2 points indicating that for each of the variables around 68% percent of the samples had a change score within 2 ladder steps above or below the mean. The only exception was 'Living skills' for the unplanned group as it had a standard deviation of 2.7 indicating a greater degree of variability. However, the sample size of the unplanned group (i.e. 15) must be taken into account if interpreting this final point.

4. Learning and recommendations

4.1 Consistency

4.1.1 For organisations

Organisations should expect a variance of between 1 and 3. Where the variance is higher, interpretation of wording may be an issue. There is a need for organisations to develop a shared understanding of how the wording should be interpreted.

Recommendation - Implementing formal training throughout the organisation improves consistency, and where practicable this should be implemented

There is some indication that consistency of use of the Star might improve with practice

Recommendation - Time should be allowed time for implementation backed up by the enabling of a shared understanding via team meetings and frequent discussion during the early months

There was some indication that conservative scores tend to be more consistent and formal training might result in more conservative scores

Recommendation - to enable training

Using the ladder results in more consistency

Recommendation - to use the ladder and to regularly revisit it

4.1.2 For commissioners

These findings below can be used to support your organisations to be more consistent.

The variation between different agencies can be much higher (3-5) than within an organisation

Recommendation - this needs to be borne in mind if comparing different organisations

There is a need to be mindful that the tool has different purposes (key-working v. data collection) and that enforcing one will have implication for the other

Recommendation – commissioners need to be clear why they want organisations to use the tool

4.1.3 For the Star

Some scales of the Star have a higher variation than others – this could be attributable to the wording.

Recommendation - to consider whether there is any scope to simplify the wording.

Different Stars have been suggested for particular client groups as a way of improving the consistency

Recommendation – to further explore the viability of introducing variations of the Star applicable to specific client groups.

4.2 Correlation to hard outcomes

A higher percentage than anticipated of the selected cases had 'missing' data where a score had not been recorded for a particular scale of the Star.

Recommendation

That organisations are encouraged to complete all scales of the Star.

The analysis produced a result that is counter-intuitive, and consequently, it might be useful to revisit the Link data when there has been an opportunity to disseminate good practice in completing the Outcomes Star. However, there is currently no evidence that there is a statistical linkage between the progress recorded by the Outcomes Star (the 'soft' outcomes) and that recorded by the Link database (the 'hard' outcomes).

Recommendation

That consideration should be given to revisiting the Link data in the future, when organisations have had the opportunity to train staff to fully complete all scales of the Outcomes Star.