

*Notes from the children and young people's mental health field: Using movement towards goals as a potential indicator of service change and quality improvement*

Jenna Jacob, PhD\*<sup>1, 3</sup>, Julian Edbrooke-Childs, PhD<sup>1, 2, 3</sup> Luís Costa da Silva, PhD<sup>1,3</sup>

& Duncan Law, D.Clin.Psy<sup>1, 2, 4</sup>

<sup>1</sup>Child Outcomes Research Consortium, Anna Freud Centre, London, UK

<sup>2</sup>Evidence Based Practice Unit, Anna Freud Centre, London, UK

<sup>3</sup>Department of Clinical, Educational and Health Psychology, UCL, London, UK

<sup>4</sup>MindMonkey Associates Ltd, London, UK

[Jenna.jacob@annafreud.org](mailto:Jenna.jacob@annafreud.org) ORCID ID 0000-0003-1006-1547

[Julian.childs@annafreud.org](mailto:Julian.childs@annafreud.org) ORCID ID 0000-0003-0401-4058

[Luis.costa@annafreud.org](mailto:Luis.costa@annafreud.org) ORCID ID 0000-0001-8955-8024

[Duncan.law@annafreud.org](mailto:Duncan.law@annafreud.org) ORCID ID 0000-0001-7873-2031

\*Corresponding author: [jenna.jacob@annafreud.org](mailto:jenna.jacob@annafreud.org); CORC, Anna Freud Centre, Kantor Centre of Excellence, 4-8 Rodney Street, London, N1 9JH. Tel: 02074432225.

**Key words:** *goals, idiographic, outcomes, routine outcome monitoring, patient reported outcome measures, indicator of change, goal-based outcomes*

**Declaration:** Duncan Law developed the GBO, trains people in its use and advocates for the clinical utility of idiographic tools. Jenna Jacob and Duncan Law have produced guidance on the Goal-Based Outcome tool (GBO). Jenna Jacob, Julian Edbrooke-Childs and Luís Costa da Silva currently work as part of CORC which encourages the use of idiographic goal-based outcome measures. Duncan Law is a Director of CORC.

**Acknowledgements:** The authors would like to thank all members of CORC, its Board at the time of writing— (includes D.L.) Ashley Wyatt, Beth Ingram, Ann York, Kate Dalzell, Isobel Fleming, Kate Martin, Amy Mary Rose Herring, Mick Atkinson, Rebecca Lewis—and the CORC team at the time of writing: (includes J.J., J.E.C & L.C.S.) Kate Dalzell, Ben Ritchie, Nick Tait, Martha Reilly, Sally Marriott, Lee Atkins, Anja Teichert, Anisa Rashid, Florence Ruby and Rachel Piper.

**Funding:** The authors did not receive specific funding to conduct this research.

**Ethics:** Ethical approval for this research was not required because it involved secondary analysis of administrative data. Youth services who have contributed data to the dataset have been advised to gain consent from participants that their data could be used for routine reporting and research purposes.

**Data availability:** The data that support the findings of this study are available from the Child Outcomes Research Consortium. Restrictions apply to the availability of these data, which were used under license for this study. Data are available <https://www.corc.uk.net/media/1883/request-for-use-of-corc-dataset-27-09-2018.doc> with the permission of the Child Outcomes Research Consortium.

### **Abstract**

The aim of this paper is to report our notes from the field on using movement toward goals at an aggregate level as an inference of service effectiveness. Analysis of routinely collected data from UK youth mental health services was conducted ( $N=8,172$ , age  $M=13.8$ , 67% female, 32% male) to explore the impact of including goal-based outcome data in combined calculations of standardized measures based on the principles of reliable change ('measurable change'). Due to the broad nature of standardized measures, inferred validity becomes diluted in any team or service level aggregate analysis. To make inferences that are closer to the person's interpretation of their difficulties, we argue that Idiographic Patient Reported Outcome Measures (I-PROMs) counterbalance these limitations. This is supported by our findings. The measurable change metric is the first step towards enabling national analysis of aggregated I-PROMs. I-PROMs, supplemented by standardized measures should be used to consider service evaluation.

### **Introduction**

Outcome measurement is a key element of tracking clients' change in health settings (Lambert 2007) and in evaluating service effectiveness (Clark et al., 2018). A variety of outcome measures are widely used across physical and mental health settings (Fleming et al., 2016; Hurn et al., 2006). Outcome measures with a clinical focus are often formed based on what clinicians feel it is important to work on. The focus is therefore usually related to symptomology and functioning (Wolpert et al., 2014). These measures tend to be normative and standardized, usually in the form of brief questionnaires consisting of established, fixed items which are conducive to making comparisons across groups (Barkham et al., 2001; Barkham et al., 2010; Lutz et al., 2005). These comparisons across groups are made by aggregating the scores from questionnaires at a team or service level, to derive means and to analyze the data at a group level. Analysis of aggregate data derived from normed, standardized or nomothetic, outcome measures can inform service delivery through benchmarking between services and sectors (Clark et al 2007; Fleming, et al., 2016; Jacob et al., 2017<sup>a</sup>, Rodgers, 2017). Normed, standardized outcome measures that provide feedback to therapists in clinical settings on progress, or lack of it, have been demonstrated as useful in adult populations to facilitate better outcomes, particularly when the client is veering off of an expected trajectory of change (Boswell et al., 2015; Carlier et al., 2012; Gondek et al., 2016; Green & Latchford, 2012; Lambert & Shimokawa, 2011; Sapyta et al., 2005; Worthen & Lambert, 2007). Whilst there is some new and emerging evidence of effectiveness of using these measures in child and youth settings (for example, Cooper et al., 2019), generally, the evidence is substantially less robust and developed than in adult settings (Bergman et al., 2018; Haynes et al., 2009).

The focus on the use of data derived from normed, standardized outcome measures to assess change in cohorts of clients is partly due to these measures having widely available norms and reasonable psychometric properties (Green, 2016). However, there is also a dominant perspective that symptoms are the most important area of change to

track progress in (Duncan et al., 2011). While this type of measurement provides clinicians and others with a good sense of the progress of groups of clients based on what appears to be valid and reliable measurement, there are also challenges with this. For example, while some of these fixed items are likely to be relevant to some clients, varying numbers of items will not be relevant to all clients. This may be particularly true in populations that are culturally and ethnically diverse. Consequently, the validity of the measures for individuals within the cohort are overestimated (Haynes et al., 2009). Similarly, as normed, standardized tools track change using professionally-defined items (that is, clinically determined), they may not capture the issues most relevant to the client and therefore may underestimate the magnitude of change experienced by the client themselves (Alves et al., 2013; Arnold & Jensen 1998, Arnold et al., 2003).

A recent systematic review of the types of measures used to track adolescent depression and anxiety found few studies assessing I-PROMs areas such as personal growth, empowerment and progress towards goals (Krause et al., 2019). However, evidence for change is shifting, to consider Idiographic Patient Reported Outcome Measures (I-PROMs) as better, or at least complementary, indicators of change due to the alignment with the individual's perception of relevant change (Arnold & Jensen 1998, Di Malta et al., 2019; Harper Shehadeh et al., 2019, Sales & Alves 2012). There have also been recent calls for more understanding of I-PROMs (Cuijpers, 2019). I-PROMs increase the relevance to the individual and therefore diverge between clients. Although the items are individual, these outcome measures can vary from those that are standardized in their delivery and scaling (Haynes, et al., 2009), such as the Goal-based Outcomes tool (GBO; Law, 2011; 2019), Goal Attainment Scaling (GAS; Kiresuk & Sherman, 1968), Goals Form (Cooper, 2015), Youth Top Problems (Weisz et al., 2011), Target Complaints (Battle et al., 1966), Psychological Outcome Profiles (PSYCHLOPS; Ashworth et al., 2004 and PSYCHLOPS Kids; Godfrey et al., 2019) and the Personal Questionnaire (PQ; Elliott et al., 2016), to those that are completely unstandardized and entirely client-defined in terms of item development, scaling and delivery. These are likely to be ad hoc, individualized methods of tracking change, such as using narrative approaches to data collection, which includes gathering verbatim feedback from individuals. Whilst considered important, there are numerous reasons that truly idiographic measures are not practical for reporting on service effectiveness at an aggregate level. The most pertinent reason is likely to be the resource implications associated with the analysis of this type of measurement.

I-PROMs such as the GBO tool (Law 2011; 2019), GAS (Kiresuk & Sherman, 1968) and the Goals Form (Cooper, 2015) are tools that facilitate discussion with clients about their hopes for the outcomes of a therapeutic encounter and expediate shared-decision making about therapy goals. These, in turn, impact the development of a stronger therapeutic alliance and thus may increase the likelihood of improved therapy outcomes and satisfaction (Cooper & Law, 2018; Jacob et al., 2017<sup>b</sup>; Law et al., 2019; Tryon et al., 2018). To date, the theory of using goal-based measures to improve outcomes is untested in robust clinical trials (Bergman et al., 2018). However, research to test the strength of this hypothesis is currently underway in north America (data should start to emerge from this trail in 2021;

## **RUNNING HEAD: GOALS AS A POTENTIAL INDICATOR OF SERVICE CHANGE**

Good Clinical Practice Network, 2020). I-PROMs, such as goal-based tools, have been found to have good face validity and allow individuals to easily access and review progress, providing a client-accepted and client-centred method of tracking change (Badham, 2011; Law & Jacob, 2015; Levack et al., 2015; Moran et al., 2012; Toto et al., 2015).

Evidence is emerging, which supports the theoretical links between measuring progress using a goal-based tool such as the GBO, and increased sensitivity to change for mental health and wellbeing outcomes (Tryon et al., 2018). Research suggests that I-PROMs cover areas of inquiry not otherwise explored by normed measures of outcome, which gives them a unique place alongside more symptom-focused measures (Jacob et al., 2017<sup>c</sup>; Sales & Alves, 2016). The main benefit of I-PROMs is their individualized nature. However, this is also a challenge, where the concern is that on an aggregate level, the very personalized nature of the measures is lost (Elliott et al., 2016; Maggin & Chafouleas, 2013). The challenge then is how to make inferences about service effectiveness based on data from such individually focused measures, particularly where there are requirements to centrally evidence this (for example, some national programmes across England; CYP IAPT, see Fonagy et al., 2017). Service providers are encouraged to use aggregate outcome information to inform discussions amongst teams and with commissioners about service provision (Fleming et al., 2016). While some evidence for validity has been found for goal-based measures (Edbrooke-Childs et al., 2015), the most robust way to analyze these cohort level data have not been widely published (Lloyd et al., 2019). Services across England currently report on the reliable change of standardized measures, such as the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) and the Revised Child Anxiety and Depression Scale (RCADS; Chorpita et al., 2005) to assess symptom and impact changes. Experience from the Child Outcomes Research Consortium, a membership organisation that collects and uses evidence to improve children and young people's mental health and wellbeing (CORC, 2019), has demonstrated that aggregate analysis of I-PROMs has, to date, been focused on mean change scores, including funnel plots (Wolpert et al., 2014), with each outcome measure reported on separately.

There has been a lack of evidence of clinicians and others using I-PROMs data to trigger quality improvement initiatives or to track changes in the quality of service provision overall. By this we mean making changes to a service, based on findings from outcome information and additional operational data about the service. These changes may involve physical alterations to the clinical setting, making decisions about the type or amount of therapies provided, or the impact of clinician training programmes. Our experience is that the principle of setting and tracking goals is fundamental at this moment in time. Service leads are not including goal change in their decision-making processes, and commissioning targets are related to the setting of goals, rather than to goal attainment. Goal-based measures are commonly perceived as most useful at the individual level and any targets set are usually related to the use of goals with youths rather than movement on the scale. The aggregate data from goal-based measures are used to assess the types of goals youths set in services, to focus work, identify gaps in the service and provide training. However, we want to consider whether they can be routinely used at the aggregate level to also consider overall outcome and service

effectiveness. This, in part, suggests that the dominant perspective may be that I-PROMs are more pertinent to tracking outcomes of the individual, than as a potential indicator of service level quality improvement. However, I-PROMS are now being used, for the first time in youth mental health settings, to assess service level change. The current metric being piloted by NHS England and NHS Improvement, within sites, to measure aggregate improvement in young people accessing child and adolescent mental health services in England is called 'measurable change' (GBO calculation inception paper: Edbrooke-Childs et al., 2015). This incorporates the principles of the reliable change index (Jacobson & Truax, 1992) of standardized outcome measures and progress towards goals. This calculation explores change which is more than would be expected solely due to measurement error. To the best of our knowledge, no study has examined the impact of combining progress towards goals with reliable change in standardized measures, in order to assess aggregate change, to consider service quality, or to explore sensitivity to change. The aim of this paper is to report our notes from the field on using movement toward goals at an aggregate level as an inference of mental health and wellbeing outcome and service effectiveness.

### **Measurable change**

The GBO allows youths to set up to three goals, tracking the progress along the way or at the end of an intervention or therapeutic contact. The rating is made on a scale from zero to ten where 'zero' means the goal has not been met in any way and 'ten' means the goal has been completely met. Edbrooke-Childs and colleagues (2015) consider how best to allow standardized analysis of cohort data that is derived from client-defined goals by calculating scores using the principles of the reliable change index (Jacobson & Truax, 1992). The GBO 'meaningful change' calculation of three or more points on the scale, was based on the original calculation by Edbrooke-Childs and colleagues (2015) and on consultations with clinicians where it was agreed that three points would represent a clinically meaningful improvement in the lives of young people.

We conducted a secondary analysis of routinely collected data from child and adolescent mental health services across the UK to further explore the impact of including goal-based outcome data in combined reliable change calculations for standardized outcome measures. The dataset used to gather information from the field is a routinely collected dataset consisting of data pertaining to children and youths (0-25 years old) who have attended a mental health service in the UK. Whilst most contributing services work with children and youths up to age 18, some do extended work up to age 25. For those who work with babies indicated as 0 years, this involves work with parents. Demographics and outcome data in the dataset were collated by CORC. Ethical approval for this research was not required because it involved secondary analysis of administrative data (NHS Health Research Authority, 2018). Youth services who have contributed data to the dataset were advised to gain prior consent from participants that their data could be used for routine reporting and research purposes.

The national reporting criteria were followed, whereby outcomes are calculated on closed cases with a minimum of three contact points and at least one paired case on a child-reported measure. An event refers to a point of contact between the client and the professional; most likely to be face-to-face meeting, but also encompassing telephone activity and outcome measurement if outside of a treatment session. The dataset with any closed cases with three or more recorded events consisted of 23,373 cases.

Two analyses were conducted using the principles of the reliable change index (Jacobson & Truax, 1992), extended to encompass several measures, and thus termed measurable change. One subset of the data comprised cases where GBO data were not available and the second analysis was conducted on the subset of the data, also including GBO data, where this was available. In both analyses, for a case to be classed as measurably improved, at least one of any of the scales had to demonstrate measurable improvement and no scale could have demonstrated measurable deterioration. The first step of this analysis was a calculation of reliable change for a group of standardized, normed measures in the dataset (see appendix for a list of measures) and the second was the same calculation, but inclusive of the child-rated GBO data. Our GBO analysis utilized change for each goal separately.

The subset of cases with three or more recorded events and at least one paired case on a child-reported measure excluding goals data consisted of 7,414 cases, with youths in this sample having on average 4.73 paired outcome measures recorded. The subset of cases with three or more recorded events with at least one paired case on a child-reported measure, which also included goals data consisted of 8,172 cases, with youths in this sample having on average 4.80 paired outcome measures recorded. Table 1 below displays the sample sizes, age ranges and means, gender proportions and ethnicity proportions for each sample.

*[Table 1 to go about here]*

Table 2 below displays the results from the measurable change calculation, both excluding and including goal data. As the table shows, the inclusion of goal data in the calculation increases the proportion of youths reporting overall measurable improvement from 45% to 53% and reduces the proportion of youths reporting no overall measurable change from 45% to 37%. There was a significant association between overall measurable change and goal inclusion,  $\chi^2(2) = 65.29, p < .001$ . A pairwise z-test post hoc analysis with a Bonferroni correction showed significant differences between the goal sample and the non-goal sample on the proportions of overall measurable improvement and no overall measurable change. The pairwise z-test post hoc analysis was conducted with a corrected p-value given by the formula:  $p/\text{number of rows} \times \text{number of columns}$ , (i.e.  $0.05/3 \times 2$ ), where columns and rows refer to the contingency table. The resulting p-value was approximately 0.008. This is the method suggested by Maxime Hervé for the R Package 'RVAideMemoire'.

*[Table 2 to go about here]*

## **Reflections**

The aim of this paper was to report our notes from the field on exploring the data from an I-PROM at an aggregate level, as an inference of sensitivity to change of the measurement of outcome and service effectiveness. We conducted secondary analyses of routinely collected data from UK child and adolescent mental health services to further explore the premise of combining idiographic aggregate data with standardized aggregate data analysis. It is important to note that such analyses should be used alongside other sources of information about the service when used for evaluation purposes. Biases in our dataset include the high proportion of missing data and a non-ethnically diverse sample. By highlighting the key challenges, we risk providing further reasons against the use of I-PROMs for those who do not want to persevere with finding ways to meet the statistical challenges they bring.

One of the biggest challenges of all types of outcome measures is inferential validity (Cronbach & Meehl, 1955; Loevinger, 1957). Inferential validity means that we are only able to measure a conjecture of difficulties: an interpretation of a person's difficulties, or an inferred representation of their mental state. Because mental health is not tangible, this is always the case despite measurement type. We argue that this inferred validity, or whether a measure can be used to make a specific accurate inference, becomes diluted in the aggregate analysis of broad measures, due to the general nature of such measures. To make a more accurate inference, or at least one that is closer to the person's interpretation of their experiences or difficulties, we argue that I-PROMs go some way to counterbalance these limitations, due to their individualized nature. It is also argued that conceptually, goal setting and tracking is set apart from standardized outcome measurement because goal tracking can be used as a method of self-regulation (Harkin et al., 2016), but also because of the locus of control and motivational aspects of goal-striving and achievement (e.g. Karoly, 1993; Elliot & Harackiewicz, 1994; Locke, 2000) which are not necessarily applicable for other measures. That is, there is more motivation to achieve a personal goal than to reach a specific score on items on a standardized outcome measure.

Goal-based measures are both client-defined and idiographic (Cooper & Law, 2018) in that what is measured: the goals chosen, are unique to the client, shaped by their context and motivations with which they enter therapy. These artefacts of goal-based outcome tools present particular statistical challenges when analyzing aggregated data: everyone's goals are different, and rating of goal progress is related to the clients' internal constructs of change. Indeed, the content of the goals set by youths and parents as part of tracking therapeutic outcomes has been demonstrated as diverse (Jacob et al., 2016; McCarthy & McDevitt, 2018). Specific challenges of aggregate I-PROM data analysis relate to the individual nature of goals and other I-PROMs and whether it is appropriate to analyze progress at a cohort level (Elliott et al., 2016; Maggin & Chafouleas, 2013). Further, when employing traditional statistical techniques such as reliable change to any data, there is an assumption that the measurement items remain static (Bollen & Diamantopoulos, 2017). This may not be the case when exploring I-PROMs including goals, which are likely to differ between participants and may change over time. It may be argued that aggregate idiographic measurement is more predisposed to



interpretational confounding if diverse items are aggregated up in calculations. Further, there is also an argument that unexplained variance may be the result of applying traditional techniques to I-PROMs that these techniques have not been developed for use with (Wilcox et al., 2008). Additionally, research has demonstrated that the misspecification of the correct analysis model can in turn lead to inaccurate conclusions about the nature of the structural relationships between constructs (Podsakoff et al., 2013).

Yet, higher levels of change in aggregate GBO have been found compared to levels of change in normed, standardized measures and change in goals showed stronger associations with change in clinician-reported functioning and service user satisfaction at the end of care. This coupled with some evidence for good internal consistency may suggest that, in fact, data from goal-based measures can be analyzed effectively at a service or cohort level (Edbrooke-Childs et al., 2015). Yet, the assumption when analyzing aggregate data is that the components are measuring the same construct and so there are challenges to the exploration of internal consistency when looking at I-PROMs (Meier, 2008). This makes this finding especially interesting, which warrants further exploration and consideration. The suggestion for the use of aggregate goal analysis by Edbrooke-Childs and colleagues (2015) is to focus on the direction and magnitude of change in relation to the goal rather than focusing on the individualized 'item' (goal) that is being measured.

Our analysis found that the inclusion of the goal-based data into the measurable change calculation demonstrated an increase in the overall level of measurable improvement and a reduction in the no measurable change category reported for youths. Further, we found a significant association between measurable change and goal inclusion, where there were significant differences between the goal sample and the non-goal sample on the proportions of overall measurable improvement and no overall measurable change. These results also demonstrate higher levels of improvement and lower levels of deterioration and no change, compared to a recent meta-analysis of reliable change in youth services. Excluding goal data, the individual-level change on self-report measures was found to be 38% reliable improvement, 44% no reliable change, and 6% reliable deterioration (Bear et al., 2020). This supports the argument for the inclusion of I-PROMs in such calculations at an aggregate level. Because lower levels of measurable improvement were demonstrated when goal data was not included in analysis, it suggests that important improvements may be missed with the exclusion of I-PROMs data analysis at this level. Further, our findings may support our argument that I-PROMs increase inferential validity, because more young people demonstrate measurable improvement with the use of the goals measure than others. However, there may be an argument to say that the more measures that are included in the measurable change analysis, the more likely it is that measurable change will be detected on at least one. To the contrary, the key argument here is that it is not necessarily a case of adding more of any type of measure, not least because the mean number of measures in both of our analyses were similar. Research suggests that goal-based measures demonstrate more positive change than symptom-focused measures, aligning more to symptoms of

functioning (Edbrooke-Childs et al., 2015). Therefore, fundamentally, we can expect that more measurable change will be detected by the inclusion specifically of goal-based measures. Consequently, to increase inferred validity, it is essential to include in these calculations measures that track something meaningful to the individual. The more precise the items are to the individual's needs, the better, and with effective interventions with clients, the more measurable change will be reflected. There is an argument to say that whether inferred validity is increased or not with the inclusion of goal data is dependent upon the focus of change, where if the key area of change is symptomology, it may be the case that including goals dilutes this due to the varied foci, whereas focusing on all types of change increases it. Nevertheless, this move towards exploring a range of important areas of change, and the trial of the measurable change metric is highly encouraging and provides the first step towards enabling the analysis of I-PROMs at an aggregate level.

However, there are still further challenges of aggregating ratings on goal-based measures. By the very context in which clients are asked to choose goals to work on in therapy, they tend to choose goals that are already at or close to the lowest point on the rating scale. If, throughout the course of therapy, no progress towards goals is made, or a deterioration occurs, there is little or nowhere for the ratings to move down the scale, thus the ratings 'bottom out'. This phenomenon leads to false underestimations of the cases that may have otherwise shown deterioration – whether this is indicated as measurable (or reliable) or not, this is an issue which would be true even in change score analysis on an aggregate level. This challenge is mitigated to some extent by goal-based measures that have both negative and positive ends of the rating scale (for example, GAS, Kiresuk & Sherman, 1968). However, even this 5-point scale, leaves only a relatively small slot in which both deterioration and improvement may be measured.

There has not yet been an exploration of whether scores across goals are equitable. This is a challenge transversely with all types of scored measures; the issue being that rating something on a scale is subjective, as are mental health experiences. Additional challenges include the short-form nature of recorded goals, which are taken from the wider context of the clinical conversation. It is not always possible to collect the content of goals for aggregation due to limitations of sharing sensitive data for analysis, for example with an organisation such as CORC or for national reporting by NHS Digital in the UK. Therefore, one must be mindful that what is represented in goal form is a shorthand version of what is being attained. The truncation process may mean that interpretations of what the true goal is and what the data are saying may vary. However, this may also be true of responding to a likert scale on a normed, standardized questionnaire.

One attempt to mitigate the complexities of statistical analysis of goal-based measures is to categorize the goals clients choose to work on, by creating 'taxonomies' of goals (adults: Little, 1983; Pöhlmann, 2001; Winell, 1982 and youths: Bradley et al., 2013, Duncan et al., 2019; Rupani et al., 2014 Odhammar & Carlberg, 2015;). There are different levels of goal taxonomies and goal categories, whereby 'levels' of goals may also be considered, such as 'immediate', 'life' and 'therapy' goals or 'symptom', 'functional' and 'fundamental' goals (Grey et al., 2018; Vermunt, 2018). This level

of goal categorization for analysis at the aggregate level may be useful as a less resource intensive way of working and recording goals (see also, Grey et al., 2018). There is some argument to ensure that the content of goals is captured to facilitate the data analysis and exploration process. Goal categories are considered a useful aid to decision making and resource planning amongst CORC members. In fact, categorizing goals is considered a key element of considering outcomes using goal-based measures. Some clinicians categorize goals into themes as soon as they are set, to assist with the burden of data entry (Grey et al., 2018). However, the challenge with categorical analysis of goal-based outcome data is that it may be perceived as defeating the very positive uniqueness of a truly client-defined measure. This is less problematic if the categorization happens after the goals have been set, but there are concerns that goals may be forced into categories if this is implemented at the start of an intervention. Further, the themes derived from goals set at the outset of therapy have been found to be helpful for identifying training needs and for service planning. CORC members have reflected how useful goal-based measures are in their settings. This is in relation to the flexibility of the measure, that they address areas that are not currently captured by other measures; and that youths care about what it is to be worked on when they can see it written down as a goal.

Due to the prevalence of research into normed, standardized measures, there is a potential danger of being lulled into a false sense of security that these measures are revealing some 'truth' about the clients' worlds and their experience of change. The danger being that the use of standardized measures that have been validated are widely considered to be more reliable and objective. Thus, outcomes are often presented as the reduction and restructuring of complex difficulties, into a limited number of variables. However, this is the challenge for mental health research and evaluation, because perhaps this approach is too reductionist. Further, robust clinical trials are mainly based on the collection and analysis of standardized data sources, the findings from which often lead to commissioning decisions. However, the consideration here is, what if the data from these measures and therefore the use of them in this way, are also ultimately flawed? Additionally, 'good' outcomes are considered those that can be measured by widely available measures which typically consist of clinically defined items (Rodgers, 2017). This is an assumption that must be made with caution at a time when emerging research questions the validity of these standardized measures (Alves et al., 2018; Arnold & Jensen, 1998; Black et al., 2020; Rodgers, 2017) which are still subjective and flawed, as all measurement tools are. Therefore, there is an argument to continue to aggregate goal data across youths and to learn from the findings in a similar way; to enable the continued research into these measures. Further, clients are unlikely to be tracking outcomes that are meaningful to them if a general approach to measurement implementation is used. This is where all clients in a service are administered the same measures to enable more straightforward benchmarking. Therefore, there is an even more compelling argument for using I-PROMs of change. That is, whatever else is tracked in therapy, what must always be measured is whether the client themselves feels that therapy is making a difference to their lives and making a difference to the things they most hope to change. The only way to do this is to use I-PROMs

of change, despite all the challenges they bring. A refusal to measure the things that clients feel are most important just because it is a difficult and challenging thing to do is a disservice to the clients the services are commissioned to help.

### **Conclusions**

Due to the personalized nature of idiographic, I-PROMs, the analysis of the data on an aggregate level may be considered too complex. Several challenges of aggregate I-PROMs analysis have been considered here. The key element to any outcome measure is knowing that the client is moving in a positive direction. Turning this to goal-based outcomes means that the goal itself may be individual, but the client's rating of the direction and magnitude of change is as standardized, as it is in any normed, symptom-based measure. Because the focus is on the client-defined areas of importance, this may lead to increased inferred validity. Using the principles of the reliable change index in the measurable change calculation allows us to analyze I-PROMs in the same way we analyze normed, standardized measures. With all measures, it is not possible to truly measure validity due to the intangible nature of mental health and wellbeing. Due to this, there may always be false positives in the findings where the evidence of change in itself is attributable to the positive or negative effects of an intervention. Through the application of the principles of the reliable change index, we aimed to reduce the likelihood of false results due to measurement error. All measures have their flaws and limitations and we must work with the best available information that we have, which is constantly developing. The key element of the use of I-PROMs is that they are used collaboratively and in a personalized way to the individual (Di Malta et al., 2019). The considered use of I-PROMs alongside normed, standardized measures is encouraged (Alves, et al., 2018; Ashworth et al., 2019; Edbrooke-Childs, et al., 2015; Green, 2016; Sales et al., 2007; Wolpert et al., 2014). It is important that this is conducted through careful consideration of the results derived from outcome measures, as well as any additional information about the service overall. This is not an either or. The best approach is to triangulate a range of different measures of change, all with their flaws and all with their strengths: functional or symptom change, satisfaction and goal progress used together are likely the most robust way of measuring outcomes.

## References

- Alves, P. C., Sales, C. M., & Ashworth, M. (2013). Enhancing the patient involvement in outcomes: a study protocol of personalised outcome measurement in the treatment of substance misuse. *BMC Psychiatry*, 13(1), 337. <https://doi.org/10.1186/1471-244X-13-337>
- Alves, P. C. G., Sales, C. M. D., Ashworth, M., & Faísca, L. (2018). "There Are Things I Want to Say But You Do Not Ask": a Comparison Between Standardised and Individualised Evaluations in Substance Use Treatment. *International Journal of Mental Health and Addiction*, 1-14. <https://doi.org/10.1007/s11469-018-9985-6>
- Arnold, L. E., & Jensen, P. S. (1998). *Interpretive Interview with Child/adolescent and Parents*. In J. Noshpitz (Ed.), *Handbook of Child and Adolescent Psychiatry: Clinical Assessment*. John Wiley & Sons.
- Arnold, L. E., Vitiello, B., McDougle, C., Scahill, L., Shah, B., Gonzalez, N. M., ... & Cronin, P. (2003). Parent-defined target symptoms respond to risperidone in RUPP autism study: customer approach to clinical trials. *Journal of the American Academy of Child & Adolescent Psychiatry*, 42(12), 1443-1450. <https://doi.org/10.1097/00004583-200312000-00011>
- Ashworth, M., Guerra, D., & Kordowicz, M. (2019). Individualised or standardised outcome measures: a co-habitation?. *Administration and Policy in Mental Health and Mental Health Services Research*, 46(4), 425-428. <https://doi.org/10.1007/s10488-019-00928-z>
- Ashworth, M., Shepherd, M., Christey, J., Matthews, V., Wright, K., Parmentier, H., ... & Godfrey, E. (2004). A client-generated psychometric instrument: The development of 'PSYCHLOPS'. *Counselling and Psychotherapy Research*, 4(2), 27-31. <https://doi.org/10.1080/14733140412331383913>
- Badham, B. (2011). *Talking about talking therapies: thinking and planning about how to make good and accessible talking therapies available to children and young people*. YoungMinds.
- Barkham, M., Hardy, G. E., & Mellor-Clark, J. (Eds.). (2010). *Developing and delivering practice-based evidence: A guide for the psychological therapies*. John Wiley & Sons.
- Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., . . . McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*, 69(2), 184-196. <https://doi.org/10.1037/0022-006X.69.2.184>

- Battle, C. C., Imber, S. D., Hoehn-Saric, R., Stone, A. R., Nash, E. R., & Frank, J. D. (1966). Target complaints as criteria of improvement. *American Journal of Psychotherapy*, 20(1), 184-192. <https://doi.org/10.1176/appi.psychotherapy.1966.20.1.184>
- Bear, H. A., Edbrooke-Childs, J., Norton, S., Krause, K. R., & Wolpert, M. (2019). Systematic review and meta-analysis: outcomes of routine specialist mental health care for young people with depression and/or anxiety. *Journal of the American Academy of Child & Adolescent Psychiatry*, 59(7), 810-841. <https://doi.org/10.1016/j.jaac.2019.12.002>
- Bergman, H., Kornør, H., Nikolakopoulou, A., Hanssen-Bauer, K., Soares-Weiser, K., Tollefsen, T. K., & Bjørndal, A. (2018). Client feedback in psychological therapy for children and adolescents with mental health problems. *Cochrane Database of Systematic Reviews*, (8). <https://doi.org/10.1002/14651858.CD011729.pub2>
- Black, L., Mansfield, R., & Panayiotou, M. (2020). Age Appropriateness of the Self-Report Strengths and Difficulties Questionnaire. *Assessment*, 1073191120903382. <https://doi.org/10.1177/1073191120903382>
- Bollen, K. A., & Diamantopoulos, A. (2017). In defense of causal-formative indicators: A minority report. *Psychological Methods*, 22(3), 581-596. <https://doi.org/10.1037/met0000056>
- Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research*, 25(1), 6-19. <https://doi.org/10.1080/10503307.2013.817696>
- Bradley, J., Murphy, S., Fugard, A. J., Nolas, S. M., & Law, D. (2013). What kind of goals do children and young people set for themselves in therapy? Developing a goals framework using CORC data. *Child and Family Clinical Psychology Review*, 1, 8-18.
- Carlier, I. V., Meuldijk, D., Van Vliet, I. M., Van Fenema, E., Van der Wee, N. J., & Zitman, F. G. (2012). Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *Journal of Evaluation in Clinical Practice*, 18(1), 104-110. <https://doi.org/10.1111/j.1365-2753.2010.01543.x>
- Child Outcomes Research Consortium (2019) Homepage. Accessed <https://www.corc.uk.net/> on 30/10/2020.
- Chorpita, B. F., Moffitt, C. E., & Gray, J. (2005). Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. *Behaviour Research and Therapy*, 43(3), 309-322. <https://doi.org/10.1016/j.brat.2004.02.004>

- Clark, D. M., Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *The Lancet*, 391(10121), 679-686. [https://doi.org/10.1016/S0140-6736\(17\)32133-5](https://doi.org/10.1016/S0140-6736(17)32133-5)
- Cooper, M. (2015). Goals Form: Guidance on use. (Working paper). Accessed [https://www.researchgate.net/profile/Mick\\_Cooper](https://www.researchgate.net/profile/Mick_Cooper) on 30/10/2020 <https://doi.org/10.13140/RG.2.1.2576.7767>
- Cooper, M., & Law, D. (2018). Introduction. In M. Cooper & D. Law (Eds.), *Working with Goals in Psychotherapy and Counselling* (pp. 1-13). Oxford University Press.
- Cooper, M., Duncan, B., Golden, S., & Toth, K. (2019). Systematic client feedback in therapy for children with psychological difficulties: pilot cluster randomised controlled trial. *Counselling Psychology Quarterly*, 1-16. <https://doi.org/10.1080/09515070.2019.1647142>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(1), 281-302. <https://doi.org/10.1037/h0040957>
- Cuijpers, P. (2019). Targets and outcomes of psychotherapies for mental disorders: an overview. *World Psychiatry*, 18(3), 276-285. <https://doi.org/10.1002/wps.20661>
- Di Malta, G. S., Oddli, H. W., & Cooper, M. (2019). From intention to action: A mixed methods study of clients' experiences of goal-oriented practices. *Journal of Clinical Psychology*, 75(10), 1770-1789. <https://doi.org/10.1002/jclp.22821>
- Duncan, C., Cooper, M., & Baxon, D. (2019). *Analysing Goals and Expectations in Counselling with Young people (AGENCY): a PhD study protocol and preliminary findings*. Paper presented at the 2nd International Conference of Pluralistic Psychotherapy and Counselling, University of Roehampton, London.
- Duncan, B. L., Miller, S. D., & Sparks, J. A. (2011). *The heroic client: A revolutionary way to improve effectiveness through client-directed, outcome-informed therapy*. John Wiley & Sons.
- Edbrooke-Childs, J., Jacob, J., Law, D., Deighton, J., & Wolpert, M. (2015). Interpreting standardized and idiographic outcome measures in CAMHS: what does change mean and how does it relate to functioning and experience? *Child and Adolescent Mental Health*, 20(3), 142-148. <https://doi.org/10.1111/camh.12107>
- Elliot, A. J., & Harackiewicz, J. M. (1994). Approach and avoidance achievement goals: An intrinsic motivation analysis. Unpublished. doctoral dissertation.

- Elliott, R., Wagner, J., Sales, C., Rodgers, B., Alves, P., & Café, M. J. (2016). Psychometrics of the Personal Questionnaire: A client-generated outcome measure. *Psychological assessment*, 28(3), 263-278. <https://doi.org/10.1037/pas0000174>
- Fleming, I., Jones, M., Bradley, J., & Wolpert, M. (2016). Learning from a learning collaboration: The CORC approach to combining research, evaluation and practice in child mental health. *Administration and Policy in Mental Health and Mental Health Services Research*, 43(3), 297-301. <https://doi.org/10.1007/s10488-014-0592-y>
- Fonagy, P., Pugh, K., & O'Herlihy, A. (2017). *The Children and Young People's Improving Access to Psychological Therapies (CYP IAPT) Programme in England*. In D. Skuse, H. Bruce & L. Dowdney (Eds.), *Child Psychology and Psychiatry: Frameworks for Clinical Training and Practice: 3rd Edition*, (pp. 429-435). John Wiley & Sons.
- Green, D. (2016). Making the case for using personalised outcome measures to track progress in psychotherapy. *European Journal of Psychotherapy & Counselling*, 18(1), 39-57. <https://doi.org/10.1080/13642537.2015.1130075>
- Gondek, D., Edbrooke-Childs, J., Fink, E., Deighton, J., & Wolpert, M. (2016). Feedback from outcome measures and treatment effectiveness, treatment efficiency, and collaborative practice: A systematic review. *Administration and Policy in Mental Health and Mental Health Services Research*, 43(3), 325-343. <https://doi.org/10.1007/s10488-015-0710-5>
- Good Clinical Practice Network. (2020). Patient generated goals in mental health: Better outcomes? Children's Hospital of Eastern Ontario. Accessed <https://ichgcp.net/clinical-trials-registry/NCT03527914> on 30/10/2020.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry*, 38(5), 581-586. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>
- Godfrey, E., Aubrey, M., Crockford, S., Haythorne, D., Kordowicz, M., & Ashworth, M. (2019). The development and testing of PSYCHLOPS Kids: a new child-centred outcome measure. *Child and Adolescent Mental Health*, 24(1), 54-65. <https://doi.org/10.1111/camh.12271>
- Green, D., & Latchford, G. (2012). *Maximising the benefits of psychotherapy: A practice-based evidence approach*. John Wiley & Sons Inc.
- Grey, N., Byrne, S., Taylor, T., Shmueli, A., Troupp, C., Stratton, P., Sefi, A., Law, R. & Cooper, M. (2018). Goal-oriented practice across therapies. In M. Cooper & D. Law (Eds.), *Working with Goals in Psychotherapy and Counselling*. Oxford University Press, pp. 181-203.



## RUNNING HEAD: GOALS AS A POTENTIAL INDICATOR OF SERVICE CHANGE

- Harkin, B., Webb, T. L., Chang, B. P., Prestwich, A., Conner, M., Kellar, I., ... & Sheeran, P. (2016). Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. *Psychological bulletin*, 142(2), 198-299. <https://doi.org/10.1037/bul0000025>
- Harper Shehadeh, M. H., van't Hof, E., Schafer, A., van Ommeren, M., Farooq, S., Hamdani, S. U., ... & Dawson, K. (2019). Using a person-generated mental health outcome measure in large clinical trials in Kenya and Pakistan: Self-perceived problem responses in diverse communities. *Transcultural Psychiatry*, 1363461519854831. <https://doi.org/10.1177/1363461519854831>
- Haynes, S. N., Mumma, G. H., & Pinson, C. (2009). Idiographic assessment: Conceptual and psychometric foundations of individualized behavioral assessment. *Clinical Psychology Review*, 29(2), 179-191. <https://doi.org/10.1016/j.cpr.2008.12.003>
- Hurn, J., Kneebone, I., & Cropley, M. (2006). Goal setting as an outcome measure: a systematic review. *Clinical Rehabilitation*, 20(9), 756-772. <https://doi.org/10.1177/0269215506070793>
- Jacob, J., De Francesco, D., Deighton, J., Law, D., Wolpert, M., & Edbrooke-Childs, J. (2017b). Goal formulation and tracking in child mental health settings: when is it more likely and is it associated with satisfaction with care?. *European Child & Adolescent Psychiatry*, 26(7), 759-770. <https://doi.org/10.1007/s00787-016-0938-y>
- Jacob, J., Edbrooke-Childs, J., Holley, S., Law, D., & Wolpert, M. (2016). Horses for courses? A qualitative exploration of goals formulated in mental health settings by young people, parents, and clinicians. *Clinical Child Psychology and Psychiatry*, 21(2), 208-223. <https://doi.org/10.1177/1359104515577487>
- Jacob, J., Edbrooke-Childs, J., Law, D., & Wolpert, M. (2017c). Measuring what matters to patients: Using goal content to inform measure choice and development. *Clinical Child Psychology and Psychiatry*, 22(2), 170-186. <https://doi.org/10.1177/1359104515615642>
- Jacob, J., Edbrooke-Childs, J., Lloyd, C., Hayes, D., Whelan, I., Wolpert, M., & Law, D. (2018). Measuring outcomes using goals. In M. Cooper & D. Law (Eds.), *Working with Goals in Psychotherapy and Counselling* (pp. 111-137). Oxford University Press.
- Jacob, J., Napoleone, E., Zamperoni, V., Levy, L., Barnard, M., & Wolpert, M. (2017). *How can outcome data inform change? experiences from the child mental health context in great Britain, including barriers and facilitators to the collection and use of data*. In T. Tilden & B.E. Wampold (Eds.). *Routine Outcome Monitoring in Couple and Family Therapy* (pp. 261-279). Springer.

Jacobson, N. S., & Truax, P. (1992). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12-19. <https://doi.org/10.1037/10109-042>

Karoly, P. (1993). Mechanisms of self-regulation: A systems view. *Annual review of psychology*, 44(1), 23-52. <https://doi.org/10.1146/annurev.ps.44.020193.000323>

Kiresuk, T. J., & Sherman, R. E. (1968). Goal attainment scaling: A general method for evaluating comprehensive community mental health programs. *Community mental health journal*, 4(6), 443-453. <https://doi.org/10.1007/BF01530764>

Krause, K. R., Bear, H. A., Edbrooke-Childs, J., & Wolpert, M. (2018). What outcomes count? A review of outcomes measured for adolescent depression between 2007 and 2017. *Journal of the American Academy of Child & Adolescent Psychiatry* 58(1) 61-71. <https://doi.org/10.1016/j.jaac.2018.07.893>

Lambert, M. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, 17(1), 1-14. <https://doi.org/10.1080/10503300601032506>

Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy*, 48(1), 72-79. <https://doi.org/10.1037/a0022238>

Law, D. (2011). *Goals and goal based outcomes (GBOs): some useful information. Version 2.0.* CAMHS Press.

Law, D (2019) The Goal-Based Outcome (GBO) tool: Guidance Notes. MindMonkey Associates.

Law, D., & Jacob, J. (2015). *Goals and Goal Based Outcomes: Some Useful Information.* 3rd Edition. CAMHS Press.

Law, M. C., Baptiste, S., Carswell, A., McColl, M. A., Polatajko, H., & Pollock, N. (1998). *Canadian occupational performance measure: COPM.* CAOT Publ. ACE.

Levack, W. M., Weatherall, M., Hay-Smith, E. J. C., Dean, S. G., McPherson, K., & Siegert, R. J. (2015). Goal setting and strategies to enhance goal pursuit for adults with acquired disability participating in rehabilitation. *Cochrane Database of Systematic Reviews*, 7. <https://doi.org/10.1002/14651858.CD009727.pub2>

Little, B. R. (1983). Personal projects: A rationale and method for investigation. *Environment and Behavior*, 15(3), 273-309. <https://doi.org/10.1177/0013916583153002>

- Lloyd, C. E. M., Duncan, C., & Cooper, M. (2019). Goal measures for psychotherapy: A systematic review of self-report, idiographic instruments. *Clinical Psychology: Science and Practice*, e12281. <https://doi.org/10.1111/cpsp.12281>
- Locke, E. A. (2000). *Motivation by goal setting*. In R. T. Golembiewski. (Ed.) *Handbook of Organizational Behavior: Second edition, revised and expanded* (pp. 43-56). Marcel Dekker, Inc.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory [Monograph No. 9]. *Psychological Reports*, 3(1), 635-694. <https://doi.org/10.2466/pr0.1957.3.3.635>
- Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., ... & Iveson, S. (2005). Predicting change for individual psychotherapy clients on the basis of their nearest neighbors. *Journal of Consulting and Clinical Psychology*, 73(5), 904-913. <https://doi.org/10.1037/0022-006X.73.5.904>
- Maggin, D. M., & Chafouleas, S. M. (2013). Introduction to the special series: Issues and advances of synthesizing single-case research. *Remedial and Special Education*, 34(1), 3-8. <https://doi.org/10.1177/0741932512466269>
- McCarthy, M., & McDevitt, S. (2018). What Do Children and Adolescents Value in Their Mental Health Recovery?. *Journal of Psychosocial Rehabilitation and Mental Health*, 5(1), 31-41. <https://doi.org/10.1007/s40737-018-0107-3>
- Meier, S. T. (2008) *Measuring Change in Counseling and Psychotherapy: 1<sup>st</sup> Edition*. New York.
- Moran, P., Kelesidi, K., Guglani, S., Davidson, S., & Ford, T. (2012). What do parents and carers think about routine outcome measures and their use? A focus group study of CAMHS attenders. *Clinical Child Psychology and Psychiatry*, 17(1), 65-79. <https://doi.org/10.1177/1359104510391859>
- NHS Health Research Authority. (2018). *Governance Arrangements for Research Ethics Committees (GAfREC)*. Accessed <http://www.hra.nhs.uk/resources/research-legislation-and-governance/governance-arrangements-for-research-ethics-committees/> on 30/10/2020.
- Odhammar, F., & Carlberg, G. (2015). Parents' and psychotherapists' goals prior to psychodynamic child psychotherapy. *European Journal of Psychotherapy & Counselling*, 17(3), 277-295. <https://doi.org/10.1080/13642537.2015.1059865>
- Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P., & Lee, J. Y. (2003). The mismeasure of man(agement) and its implications for leadership research. *The Leadership Quarterly*, 14(6), 615-656. <https://doi.org/10.1016/j.leaqua.2003.08.002>

## RUNNING HEAD: GOALS AS A POTENTIAL INDICATOR OF SERVICE CHANGE

- Pöhlmann, K. (2001). Agency- and communication-orientation in life goals: impact on goal pursuit strategies and psychological wellbeing. In P. Schmuck & K. M. Sheldon (Eds.) *Life Goals and Well-Being: Towards a Positive Psychology Of Human Striving* (pp. 68-84). Hogrefe & Huber.
- Rodgers, B. (2017). The trouble with numbers: Some fundamental flaws with using standardised outcome measures. *Psychotherapy and Politics International*, 15(3), e1423. <https://doi.org/10.1002/ppi.1423>
- Rupani, P., Cooper, M., McArthur, K., Pybis, J., Cromarty, K., Hill, A., Levesley, R., Murdoch, J. & Turner, N. (2014). The goals of young people in school-based counselling and their achievement of these goals. *Counselling and Psychotherapy Research*, 14(4), 306-314. <https://doi.org/10.1080/14733145.2013.816758>
- Sales, C., & Alves, P. C. (2012). Individualized patient-progress systems: Why we need to move towards a personalized evaluation of psychological treatments. *Canadian Psychology/Psychologie canadienne*, 53(2), 115-121. <https://doi.org/10.1037/a0028053>
- Sales, C. M., & Alves, P. C. (2016). Patient-centered assessment in psychotherapy: A review of individualized tools. *Clinical Psychology: Science and Practice*, 23(3), 265-283. <https://doi.org/10.1111/cpsp.12162>
- Sales, C., Goncalves, S., Fragoeiro, A., Noronha, S., & Elliott, R. (2007). Psychotherapists openness to routine naturalistic idiographic research?. *Mental Health and Learning Disabilities Research and Practice*, 4(2), 145-161.
- Sapyta, J., Riemer, M., & Bickman, L. (2005). Feedback to clinicians: Theory, research, and practice. *Journal of Clinical Psychology*, 61(2), 145-153. <https://doi.org/10.1002/jclp.20107>
- Toto, P. E., Skidmore, E. R., Terhorst, L., Rosen, J., & Weiner, D. K. (2015). Goal Attainment Scaling (GAS) in geriatric primary care: a feasibility study. *Archives of Gerontology and Geriatrics*, 60(1), 16-21. <https://doi.org/10.1016/j.archger.2014.10.022>
- Tryon, G. S., Birch, S. E., & Verkuilen, J. (2018). Meta-analyses of the relation of goal consensus and collaboration to psychotherapy outcome. *Psychotherapy*, 55(4), 372-383. <https://doi.org/10.1037/pst0000170>
- Tryon, G. S., (2018) Goals and Psychotherapy Research In M. Cooper & D. Law (Eds.), *Working with Goals in Psychotherapy and Counselling* (pp. 1-13). Oxford University Press.
- Vermunt, N., Harmsen, M., Elwyn, G., Westert, G., Burgers, J., Olde Rikkert, M. & Faber., M, (2018) *A three-goal model for patients with mulitmorbidty: a qualitative approach*. In: Vermunt, N. Collaborative goal setting: towards a

goal-oriented approach of shared decision-making in complex elderly care (Doctoral Thesis) Radboud University Nijmegen, Someren, The Netherlands.

Weisz, J. R., Chorpita, B. F., Frye, A., Ng, M. Y., Lau, N., Bearman, S. K., ... & Hoagwood, K. E. (2011). Youth top problems: Using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy. *Journal of Consulting and Clinical Psychology, 79*(3), 369-380.

Wilcox, J. B., Howell, R. D., & Breivik, E. (2008). Questions about formative measurement. *Journal of Business Research, 61*(12), 1219-1228. <https://doi.org/10.1016/j.jbusres.2008.01.010>

Winell, M. (1982). *Personal goals: The key to self-direction in adulthood*. In M. E. Ford & D. H. Ford (Eds.), *Humans as self-constructing living systems: putting the framework to work* (pp. 261-287). Erlbaum.

Wolpert, M., Deighton, J., De Francesco, D., Martin, P., Fonagy, P., & Ford, T. (2014). From 'reckless' to 'mindful' in the use of outcome data to inform service-level performance management: perspectives from child mental health. *BMJ Quality & Safety, 23*(4), 272-276. <http://dx.doi.org/10.1136/bmjqs-2013-002557>

Worthen, V. E., & Lambert, M. J. (2007). Outcome oriented supervision: Advantages of adding systematic client tracking to supportive consultations. *Counselling and Psychotherapy Research, 7*(1), 48-53. <https://doi.org/10.1080/14733140601140873>

**Table 1. Demographic characteristics of the total sample, paired sample (excluding goals) and paired sample (including goals)**

	<b>Total sample</b>	<b>Paired sample (excluding goals)</b>	<b>Paired sample (including goals)</b>
<b>Sample size</b>	<i>N</i> = 23,373	<i>N</i> = 7,414 (31% of total sample)	<i>N</i> = 8,172 (35% of total sample)
<b>Age</b>	Age range: 0-25 <i>M</i> = 12.43 ( <i>SD</i> = 3.84)	Age range: 0-25 <i>M</i> = 14 ( <i>SD</i> = 2.65)	Age range: 0-25 <i>M</i> = 13.8 ( <i>SD</i> = 2.84)
<b>Gender</b>			
<i>Male</i>	10,006 (43%)	2,262 (30%)	2,624 (32%)
<i>Female</i>	13,342 (57%)	5,140 (69%)	5,535 (67%)
<b>Ethnicity</b>			
<i>White British</i>	13,207 (57%)	4,360 (59%)	4,775 (58%)
<i>Any Other White     background</i>	789 (3%)	204 (3%)	219 (3%)
<i>Any Other Ethnic group</i>	446 (2%)	133 (2%)	133 (2%)

**Table 2. Child-reported measures measurable change excluding and including goal data**

	Excluding goal data ( <i>N</i> = 7,414)		Including goal data ( <i>N</i> = 8,172)	
	<i>N</i>	% [95% CI]	<i>N</i>	% [95% CI]
Overall Measurable Improvement	3327	45% [44%, 46%]	4335	53% [52%, 54%]
No Overall Measurable Change	3321	45% [44%, 46%]	3004	37% [36%, 38%]
Overall Measurable Deterioration	766	10% [9%, 12%]	833	10% [9%, 11%]

Appendix

**Table A. Measures included in the ‘measurable change’ analysis**

Name	Acronym
Strengths and Difficulties Questionnaire	SDQ
Revised Child Anxiety and Depression Scale	RCADS
Generalized Anxiety Scale - 7	GAD7
Outcomes Rating Scale/ Child Outcomes Rating Scale Patient Health Questionnaire - 9	ORS/CORs PHQ9
Impact of Events Scale	IES
Me and My School (Renamed Me and My Feelings)	MAMS
Routine Monitoring Questionnaire	RMQ
Clinical Outcomes in Routine Evaluation/Young Person Clinical Outcomes in Routine Evaluation	CORE10/YP-CORE
Goal-Based Outcomes	GBO