








Idiographic patient reported outcome measures (I-PROMs) for routine outcome monitoring in psychological therapies: Position paper

Célia M. D. Sales¹  | Mark Ashworth²  | Salma Ayis²  |
Michael Barkham³  | Julian Edbrooke-Childs⁴  |
Luís Faísca⁵  | Jenna Jacob⁴  | Dan Xu⁶ | Mick Cooper⁷ 

¹Faculty of Psychology and Education Sciences (FPCEUP), Center for Psychology at the Universidade do Porto (CPUP), University of Porto, Porto, Portugal

²School of Life Course and Population Sciences, King's College London, Guy's Campus, London, United Kingdom, SE1 1UL, UK

³Clinical and Applied Psychology Unit, Department of Psychology, University of Sheffield, Sheffield, UK

⁴Anna Freud Centre, Applied Research and Evaluation, The Kantor Centre of Excellence, 4-8 Rodney Street, London, N1 9JH, UK

⁵FCHS & Center for Research in Health Technologies and Information, Universidade do Algarve, Faro, Portugal

⁶Zhejiang University of Technology, Hangzhou, China

⁷Department of Psychology, University of Roehampton, London, UK

Correspondence

Célia M. D. Sales, Universidade do Porto, Porto, Portugal.
Email: celiasales@soutodacasa.org

Abstract

Idiographic patient-reported outcome measures (I-PROMs) are a growing set of individualized tools for use in routine outcome monitoring (ROM) in psychological therapies. This paper presents a position statement on their conceptualization, use, and analysis, based on contemporary evidence and clinical practice. Four problem-based, and seven goal-based, I-PROMs, with some evidence of psychometric evaluation and use in psychotherapy, were identified. I-PROMs may be particularly valuable to the evaluation of psychological therapies because of their clinical utility and their alignment with a patient-centered approach. However, there are several challenges for I-PROMs: how to generate items in a robust manner, their measurement model, methods for establishing their reliability and validity, and the meaning of an aggregated I-PROM score. Based on the current state of the literature, we recommend that I-PROMs are used to complement nomothetic measures. Research recommendations are also made regarding the most appropriate methods for analyzing I-PROM data.

KEYWORDS

idiographic outcomes, I-PROMs, measurement model, psychometrics, routine outcome monitoring

The debate around psychological outcome measurement has gained importance in the current context of managed health, where the assessment of outcomes is a routine component of services. Routine Outcome Monitoring (ROM), the regular assessment of patient outcomes to capture progress during the course of therapy, has become a quality tool for measuring the clinical response of patients. Exemplars of ROM include the session-by-session use of mandated outcome measures in NHS England's Improving Access to Psychological Therapies program (Clark, 2018), the primary mental health collaborative care model in the Netherlands (van Orden et al., 2009) and the TrueBlue model in Australia (Morgan et al., 2013). At the service level, the aggregation of patient scores derived from ROM has the potential of providing evidence on the outcomes of different services and treatments, thereby informing service commissioning and policy guidelines. At the clinical level, ROM has been used to inform individual patient clinical progress. Feedback on patient progress provided by ROM systems has the potential to improve treatment outcomes and optimize clinical decision-making (Lambert et al., 2018).

The assessment tools used in ROM, Patient Reported Outcome Measures (PROMs), are typically nomothetic. These are standardized questionnaires that measure patients' self-reported experiences on universal indicators of psychological health. Such questionnaires are typically well-evidenced in psychometric terms, brief, acceptable to patients, and have items that are broad enough to capture a gamut of difficulties and experiences. This allows for population-level comparisons (Barkham et al., 2001, 2010; Green, 2016; Lutz et al., 2005). The procedure for administering nomothetic PROMs to patients locates their scores against norms derived from clinical and nonclinical populations. Change in scorings of nomothetic PROMs at different points of treatment are taken as indicators of therapeutic change, and inferences can be made about the effect of the intervention.

Despite the value of nomothetic PROMs, it is widely recognized that each person receiving a psychological therapy or intervention presents with a unique configuration of characteristics, problems, strengths, preferences, and circumstances (e.g., Cooper & McLeod, 2011). On this basis, it has been argued that individualized methods are needed for assessing therapeutic change (Evans, 2012; Piccirillo & Rodebaugh, 2019). *Idiographic PROMs* (I-PROMs), or "Patient-Generated Outcome Measures", are outcome tools that involve patients in devising their own assessment criteria, reflecting the areas they need help for and wish to change (Sales, 2017). On I-PROMs, free-text items are generated by patients, and then scored on such dimensions as "intensity" or "goal progress". This allows for the computation of change scores.

While nomothetic PROMs are grounded in the positivist assumptions of classical test theory (that a "true" score exists, along a latent, objectively "real" dimension), I-PROMs are more closely aligned with constructivist, constructionist, and phenomenological thinking (Ashworth et al., 2019). This emphasizes, and prioritizes, the uniqueness of individual experiences, perceptions, and constructions. Here, patients' problems are not seen primarily as expressions of—and reducible down to—general, pre-definable syndromes, quantifiable along numerical dimensions. Rather, their manifestation in the conscious, sense-making lived-experiencing of patients—in all their difference and diversity—is, itself, considered ontologically legitimate.

This paper aims to give a position statement on the use of I-PROMs for ROM in psychological therapies. The authors are a group of researchers that have developed, and conducted research on, I-PROMs, including experts in psychotherapy research and psychometrics. We provide a description of I-PROMs and the principal measures available, discuss their strengths and limitations for the monitoring of outcomes in real clinical settings, propose criteria for assessing their psychometric properties, and consider implications for policy and practice. Through this position paper, we aim to outline new directions for practice with, and research on, I-PROMs; present expert recommendations on the key methodological challenges that need to be addressed over coming years—if the I-PROMs field is to mature methodologically; and guide current usage of I-PROMs in practice and on the analysis of I-PROMs data.

1 | I-PROMS IN PSYCHOLOGICAL THERAPIES: A NARRATIVE REVIEW

There are two major types of I-PROMs (Lloyd et al., 2019; Sales & Alves, 2016): *Problem-based I-PROMs* invite patients to indicate difficulties or concerns to be addressed during treatment; *Goal-based I-PROMs* invite patients to specify the objectives they want to achieve. Tables 1 and 2 present a review of available problem-based and goal-

based I-PROMs, respectively. The measures identified in these tables are based on recent systematic reviews (Lloyd et al., 2019; Sales & Alves, 2016), with additional searches conducted for more recent measures (up to January 2021). These additional searches were conducted using key term searches in Google and Google Scholar (including synonyms for, and combinations of, “goal-based”, “problem-based”, “outcome measures”, “patient-reported”). Eligibility criteria for inclusion in the narrative review were evidence of (a) use in psychological therapies, and (b) psychometric evaluation.

Four problem-based I-PROMs met criteria for inclusion: two developed for use with adults only, and three for young people and adult population. These I-PROMs involve two main processes for gathering individualized items: the self-report method, where patients are invited to write their concerns in a pen-and-paper format (e.g., Psychological Outcome Profiles, PSYCHLOPS; Ashworth et al., 2005); and open-ended interviews, where patients are asked, in a dialogue, to talk about their problems (e.g., Simplified Personal Questionnaire, PQ; Elliott et al., 2016).

Seven goal-based I-PROMs met criteria for inclusion (six developed for adults and one developed for work with children and young people). However, one of these, the Motivational Structure Questionnaire, can be considered a family of several discrete measures (see Cox & Klinger, 2021; Klinger & Cox, 2011). Lloyd et al. (2019) categorized these goal-based I-PROMs into three types. First are those multidimensional tools, developed primarily for psychological research, that invite patients to establish goals through structured procedures, and then to rate them on a range of dimensions (Motivational Structure Questionnaire, Personal Project Analysis, Striving List and Striving Assessment Scales). The second type, consisting of just one tool—albeit the most widely used and cited goal-based I-PROM—which is Goal Attainment Scaling (Kiresuk & Sherman, 1968). Like the first type, this has a relatively in-depth goal setting process but, uniquely, invites patients to consider and set their own expected levels of outcomes. The third type comprises brief rating forms that have a relatively simple goal setting process for use with children and young people, and just one dimension for rating goal progress: the Goal Based Outcomes (GBO) tool (Law & Jacob, 2015) and the Goals Form (Cooper & Xu, 2021). These latter forms have been developed for informing clinical practice, but also specifically for use in ROM.

1.1 | What do I-PROMs offer to ROM?

1.1.1 | Clinical utility

The creation of an I-PROM can closely parallel—or be an integral part of—clinical assessment and interviews (Greenhalgh et al., 2018). It encourages patients, both adults and young people, to express their own views, involving them directly in the establishment of the evaluation criteria of their own treatment, and tailoring assessment to the relevant and meaningful aspects of each patient's life. Research indicates that patients, even with the same diagnosis, may have very different expectations of therapy (Grosse & Grawe, 2002; Rajkarnikar, 2009). For instance, Krause et al. (2020) found different outcome priority profiles for young people with experience of depression. Improvements in mood and capacity for hedonia were a priority for all participants. However other outcomes such as learning coping skills, processing experiences, and functioning were important to differing degrees across profiles (Krause et al., 2020). Tailoring assessment using I-PROMs may contribute to the enhancement of treatment goals agreement (e.g., Sales et al., 2019), leading to the development of the therapeutic alliance (Bordin, 1979).

The process of goal- or problem-identification, in itself, has been indicated as beneficial both by adult patients and their clinicians. It is seen as providing an opportunity for self-reflection, helping patients to think more thoroughly about their difficulties and the impact these have on their lives, and leading to disclosures to the clinician that might not otherwise be made (Alves et al., 2016; Antunes et al., 2020; Di Malta et al., 2019; Greenhalgh et al., 2018; Guerra et al., 2018; Jensen-Doss et al., 2018). For both young people and adults, goal setting has been associated with increased satisfaction with care (Jacob et al., 2017; Mintz & Kiesler, 1982), as well as with increased therapeutic retention (Cairns et al., 2019). This may be because I-PROMs allow for the faster identification of

TABLE 1
 Problem-based I-PROMS

| Name (key reference[s], website) | Brief description (inc. population, item generation, item rating, assessment, scoring). Variations. | Usage to date (who by, included in monitoring systems) | Clinical utility/face validity (acceptability, clinical usefulness, feasibility) | Psychometric evidence (reliability [internal, test-retest], convergent validity, divergent validity, sensitivity to change) |
|---|--|--|---|---|
| Simplified Personal Questionnaire (PQ, Elliott et al., 1999) | PQ is generated in a semi-structured interview, where the client reports the problems that motivated him or her to seek therapy. This results in a list of items in the patient's own words (e.g., "My son does not talk to me anymore") that are rated for the degree of bother (from 1—not at all to 7—maximum possible) and for duration (from 1—less than a month to 7—more than 10 years) | The PQ is available in English, Portuguese and Spanish. | Therapists find that PQ provides useful information for establishing treatment goals, in individual and family formats; PQ is useful for monitoring progress in specific personal domains; it warns about emerging problems; it supports ongoing clinical decision making; allows saving time/number of sessions, and it helps with writing clinical reports for supervision and administrative purposes (Sales et al., 2007). It helps patients thinking about their difficulties (Antunes et al., 2020). However, the generation of PQ items is time-consuming. | Internal reliability: demonstrated as acceptable. Test-retest reliability: overall value good. Convergent validity: the PQ correlated with several outcome measures. |
| | Clients are free to add or remove items on every application Age and population: 8 years old and above; psychiatric and counselling settings, culturally diverse population | Use in ROM: PQ is included in the Individualized Patient-Progress System (IPPS) and CORE-NET | | Sensitivity to change: the pre-post standardized differences of the mean (Cohen's d) were large; on a session-to-session basis, these effect sizes were small (Elliott et al., 2016) |
| Psychological Outcome Profiles (PSYCHLOPS; Ashworth et al., 2004) www.psychlops.org.uk | PSYCHLOPS is a self-complete, one page, patient generated outcome measure. It consists of | Used as a patient-centred mental health outcome measure for talking therapies in community, primary care settings and in | Therapists report that PSYCHLOPS increase counselling engagement with patient priorities; it is culturally sensitive; it warns | Internal reliability: Cronbach's alpha is demonstrated as acceptable. |

TABLE 1 (Continued)

| Name (key reference[s], website) | Brief description (inc. population, item generation, item rating, assessment, scoring). Variations. | Usage to date (who by, included in monitoring systems) | Clinical utility/face validity (acceptability, clinical usefulness, feasibility) | Psychometric evidence (reliability [internal, test-retest], convergent validity, divergent validity, sensitivity to change) |
|----------------------------------|--|---|---|---|
| | four questions in which the client is asked: | patients with substance dependency problems. | about emerging problems; it supports ongoing clinical decision making. However, not all patients can self-complete PSYCHLOPS (Alves et al., 2020; Ashworth et al., 2005). | Test-retest reliability: ICC = 0.79 (95% CI: 0.64 –0.88). Convergent validity: Considered good with both HADS baseline scores; and HADS change scores. |
| | (a) to describe the “problem” that troubles them most, writing this in a freetext box, then scoring it (Problem 1) to add a second problem, if there is one, again writing it in a freetext box and scoring it (Problem 2) | Used globally by WHO (in their PM + and SH+ programmes working in LMICs), Médecins Sans Frontières (in Palestine) and International Rescue (with refugees in Greece). | | |
| | (b) to describe “one thing” that is hard to do because of the problem, writing it in a freetext box, then scoring it (Function 1) | | | |
| | (c) to score how they are feeling in themselves (Wellbeing 1) | | | |
| | Each question is scored 0-5, giving a maximum possible score of 20 | | | |
| | PSYCHLOPS is suitable for use by adults 17+ years. | | | |
| | PSYCHLOPS Teen is for 13–16 years old | | | |
| | PSYCHLOPS Kids is for 7–13 years old (Godfrey et al., 2019) | Currently available in 14 languages. | | |
| | | Use in ROM: Included in CORE-NET | | Sensitivity to change: Large effect sizes were demonstrated (Ashworth et al., 2009; Evans et al., 2010). |

(Continues)

TABLE 1 (Continued)

| Name (key reference[s], website) | Brief description (inc. population, item generation, item rating, assessment, scoring). Variations. | Usage to date (who by, included in monitoring systems) | Clinical utility/face validity (acceptability, clinical usefulness, feasibility) | Psychometric evidence (reliability [internal, test-retest], convergent validity, divergent validity, sensitivity to change) |
|--|---|--|--|--|
| Youth TOP Problems (TP; Weisz et al., 2011) | <p>After diagnostic assessment, young people and their parents/carers/teachers separately identify the three problems of greatest concern, written down in their own words, (e.g., "My mom and I argue a lot."). The problems are rated on a scale of 0 ["not at all"] to 10 ["very, very much"] in response to: 'How big of a problem is this for you/them'. They are prioritised based on the following questions: "Is the biggest problem right now? Which of these is giving you/ them the most trouble right now? Which one is the most important to work on?," resulting in a ranked list. The severity of the problems is rated weekly during treatment.</p> | <p>Available in English.</p> <p>Used and tested in the U.S.</p> <p>Manual for use: https://weiszlab.fas.harvard.edu/files/jweisz/files/top_problems_assessment_manual_09.11.18.pdf</p> | <p>No evidence of the clinical utility of TP could be identified.</p> | <p>Test-retest reliability correlations ranged from good to excellent (mean interval between measures: approximately 8 days).</p> <p>Convergent validity: TP correlated with several standardised instruments; correlations on subscales and total scores on both young person-reported, and parent/carer-reported measures were significant.</p> <p>Discriminant validity: Correlations between TP scores and measures of theoretically distinct constructs demonstrated that the externalising subscale on the parent-carer-reported TP was significantly correlated with the DSM Anxiety scale.</p> <p>Sensitivity to change: When compared to a standardised measure (BPC), Internalizing, externalizing, and total score reliabilities are good (parent/carer and young person) for both TP and for BPC (Weisz et al., 2011).</p> |

TABLE 1 (Continued)

| Name (key reference[s], website) | Brief description (inc. population, item generation, item rating, assessment, scoring). Variations. | Usage to date (who by, included in monitoring systems) | Clinical utility/face validity (acceptability, clinical usefulness, feasibility) | Psychometric evidence (reliability [internal, test-retest], convergent validity, divergent validity, sensitivity to change) |
|--|---|--|--|--|
| ASSERT (Tollefsen, Darrow, et al., 2020) | ASSERT has been developed in Norway for young people in primary mental health care. The therapist asks the young person 'What matters to you?' and uses their concerns to guide the intervention. At every session, young people indicate on a scale from 1 to 10 whether they are closer or further away from their self-defined concerns. A higher score indicates improvement, meaning that the young person is less bothered by this concern. | ASSERT is available in English and Norwegian | Interviews and focus groups with therapists who had used ASSERT found that it was experienced as a positive way to gain insight into young people's concerns and needs. Therapists said that ASSERT helped to focus treatment and allowed them to follow-up on concerns in a methodological way. Therapists viewed ASSERT as a way to allow young people to feel empowered about being involved in their care (Tollefsen, Darrow, et al., 2020). | Convergent validity: In an RCT, comparisons were made between ASSERT and standardised instruments (MHLIC for locus of control; Norwegian SDQ for mental health; Norwegian version of ILC for quality of life) revealed that ASSERT was associated only with locus of control. In comparisons between groups, the score [on the outcome measure] was significantly lower in the ASSERT group (M = 39.8) than in the control group (M = 42.9) (Tollefsen, Neumer, et al., 2020). |

Abbreviations: CI, confidence interval; I-PROM, Idiographic patient-reported outcome measures; ROM, routine outcome monitoring; WHO, World Health Organization.

TABLE 2
 Goal-based I-PROMs

| Name (key reference[s], website) | Brief description (inc. population, item generation, item rating, assessment, scoring). Variations. | Usage to date (who by, included in monitoring systems) | Clinical utility/face validity (acceptability, clinical usefulness, feasibility) | Psychometric evidence (reliability [internal, test-retest], convergent validity, divergent validity, sensitivity to change) |
|---|---|---|---|---|
| | | | | |
| Goal Attainment Scaling (GAS; (Kiresuk & Sherman, 1968; Kiresuk et al., 1994) | <p>Primarily used with adults. Needs to be led by trained professional. Focal issues for treatment are identified and at least three goals are chosen. An outcome indicator is then identified for each goal; and “expected”, “more/less than expected”, and “much more/less than expected” outcomes are set. This 5-point scale can then be used for follow-up rating. An average can be calculated. Several variations of basic procedure have been developed for different populations and contexts.</p> | <p>Most widely used of goal-based I-PROM. Has been applied to mental health treatments as well as a wide range of other domains, such as education, rehabilitation, drug treatment, and correction.</p> | <p>Research indicates that, “the process of setting goals [with GAS] may itself have a positive effect on treatment outcome” (Smith, 1994, p. 3); with more success in reaching goals, and greater personality adjustment (Mintz & Kiesler, 1982). Clients have also reported being more satisfied with treatment when GAS is used, and saying that they found the process “therapeutic” (Cardillo & Smith, 1994; Mintz & Kiesler, 1982). Setting and rating of goals can be time consuming, such that GAS may not be appropriate for session-by-session monitoring. Use of GAS also requires extended training period for professional (approx. 14 h).</p> | <p>Average scale intercorrelation is low (Kiresuk & Sherman, 1968; Smith, 1994). Test–retest reliability, from end of therapy to 8-week follow-up, is acceptable ($r = 0.77$, McGaghie & Menges, 1975). GAS scores show significant moderate to high correlations with other indicators of psychological health, such as the Brief Symptom Inventory (Shefler et al., 2001).</p> |
| Motivational Structure Questionnaire (MSQ), Interview Questionnaire (IntQ), Personal Concerns Inventory (PCI), Personal Offender Adaptation (PCI: OA), Personal Aspirations | <p>Family of measures, developed primarily as assessment tools for “systematic motivational counselling” (SMC) for alcohol use in adults. Clients asked to list concerns in core life areas, then reframe as goals. Goals then rated on a</p> | <p>Primarily used within SMC. Have been adapted for other problems and environments, such as the workplace, and use with adolescents in schools.</p> | <p>Client feedback on helpfulness, difficulty, and clarity suggests MSQ is feasible and acceptable to clients (Grothenrath & Schneider, 1996, cited in De Jong-Meyer, 2004). Clients suggested that the measure provided clarity around goals and</p> | <p>Internal stability for forms shows considerable variation (Klinger & Cox, 2011). Low levels of test–retest reliability on IntQ were demonstrated. MSQ has moderate convergent validity with BDI (Baumann, 2011). Criterion validity distinguishing</p> |

TABLE 2 (Continued)

| Name (key reference[s], website) and Concerns Inventory (PACI) (see Cox & Klingler, 2021; Klingler & Cox, 2011). | Brief description (inc. population, item generation, item rating, assessment, scoring). Variations. | Usage to date (who by, included in monitoring systems) | Clinical utility/face validity (acceptability, clinical usefulness, feasibility) | Psychometric evidence (reliability [internal, test-retest], convergent validity, divergent validity, sensitivity to change) |
|---|---|--|--|---|
| range of dimensions and scales, including "nearness to goal attainment"/"goal distance" (0–9). Digital Scoring algorithms available on request. IntQ is forerunner to family of measures. PCI is simplified version of MSQ, adapted to an offender population as PCI: OA. PACI focuses more directly on goals/aspirations than MSQ. | | | concerns and improved their motivation. Use of PCI associated with greater session attendance and engagement (Drieschner & Boomsma, 2008; McMurran et al., 2013). Can be time-consuming to complete (approx. 1 h for simplest versions), as well as training. PCI-OA and PACI-O acceptable to offender population but did not lead to improved outcomes (Sellen et al., 2013). | between clinical and nonclinical established (Man et al., 1998). |

| | | | | |
|---|--|--|--|--|
| Personal Project Analysis (PPA) (Little & Gee, 2007; Little, 1983). www.brianriddle.com/Topics/research/personal-projects-analysis/ | Primarily used with adults. Respondents invited to list around 10–15 of their current "projects", then appraise each one on 0–10 standardised scales, including "Likelihood of success," "Difficulty," and "progress." "Cross-impact matrix" allows respondents to rate relationship between goals. PPA can be administered through clinical interview, self-report workbook, or digitally | PPA has been used in university counselling services (Salmela-Aro, 1992), as well as group based psychoanalytic and experiential therapies (Salmela-Aro & Nurmi, 1996). No evidence of use in ROM. | No evidence. Can be time-consuming to complete assessment and subsequent rating, as well as training | Test–retest reliability shown to be moderate (Little & Coulombe, 2015). Independent correlations between each of the PPA factors, and clinical concerns like depression, have been found (Little, 2011): e.g., PPA dimension of progress had a significant negative association with depression (Dowden et al., 2001). |
|---|--|--|--|--|

(Continues)

TABLE 2 (Continued)

| Name (key reference[s], website) | Brief description (inc. population, item generation, item rating, assessment, scoring). Variations. | Usage to date (who by, included in monitoring systems) | Clinical utility/face validity (acceptability, clinical usefulness, feasibility) | Psychometric evidence (reliability [internal, test-retest], convergent validity, divergent validity, sensitivity to change) |
|---|---|--|--|---|
| Strivings List and Striving Assessment Scales (SAS) (Emmons, 1986) | Up to 15 "personal strivings" generated, each then rated on up to 15, 0–10 point dimensions, including "probability of success". | Use in both individual and CBT programs for veterans suffering from Posttraumatic Stress Disorder (Kashdan et al., 2010), as well as in motivational interventions for those with co-morbid schizophrenia and alcohol use disorders (Carey et al., 2007). No evidence of use in ROM. | No evidence. Can be time-consuming to complete assessment and subsequent rating, as well as training. | One month test–retest reliability for "probability of success" were fair to good (Emmons, 1986). Evidence of stability of strivings over time was good, with the majority of strivings remaining the same (or closely worded variations) 1 year later (Emmons, 1986). |
| Goal-based Outcomes tool (GBO; D. Law & Jacob, 2015) https://goals-in-therapy.com/tag/gbo/ | In collaboration with their parent/carer and practitioner, or separately, the GBO allows young people to set up to three goals. Progress at the start and then during therapy, or at the end of an intervention, is rated 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. | Implemented nationally by NHS England and NHS improvement as part of the children and young people's outcome metric. Available in English, Portuguese, French, Norwegian, Japanese, Welsh, Irish and Russian | Young people and their representatives have expressed positive attitudes towards the use of GBO, particularly in relation to goal setting and tracking helping to focus treatment and to empower young people to be involved in decisions about their care (Badham, 2011; Bromley & Westwood, 2013; Feltham et al., 2018; D. Law & Jacob, 2015). | Convergent validity: aggregate GBO scores were correlated to standardised parent-and-practitioner-reported instruments (SDQ, CGAS). Correlations were low to moderate. Incremental validity: Goal change explained 35% of the variance of change in functioning. Reliable change: a calculation based on the principles of the Reliable Change Index (Jacobson & Truax, 1991) resulted in a "meaningful" change score of 2.45 points or more, based on parent rated goals. Internal reliability: internal consistency between goals |

TABLE 2 (Continued)

| Name (key reference[s], website) | Brief description (inc. population, item generation, item rating, assessment, scoring). Variations. | Usage to date (who by, included in monitoring systems) | Clinical utility/face validity (acceptability, clinical usefulness, feasibility) | Psychometric evidence (reliability [internal, test-retest], convergent validity, divergent validity, sensitivity to change) |
|--|--|---|--|---|
| | | | | based on parent-reported GBO at baseline is good at both time points tested (Edbrooke-Childs et al., 2015). |
| Goals Form (Cooper & Xu, 2021). https://pluralisticpractice.com/tools-and-measures/ | Developed specifically for ROM with adults. Client and therapist identify up to seven goals for treatment; write them down to create an individualised Goals Form; then rate and re-rate on a 1 (<i>not at all achieved</i>) to 7 (<i>completely achieved</i>) Likert-type scale. Can be used session-by-session. Client can add, remove, and revise goals. Mean change scores based on pre- to post-changes on individual goals over treatment. | Used for ROM in UK-based studies of “pluralistic counselling for depression” (Cooper, 2015; Cooper & Xu, 2021). | Clients’ mean ratings of Goals Form = 3.9 and 4.2 on 1 (<i>very unhelpful</i>) to 5 (<i>very helpful</i>) scale (Cooper & Xu, 2021). Qualitative interview data indicates measure generally acceptable. Specifically developed to be brief and easy to use—suited to session-by-session monitoring. Limited training required. | Within-clients internal reliability s good or very good. Mean test–retest stability is good from assessment to session 1. Within-clients convergent validity is mixed, from low to moderate. Scores divergent from session rating scale. Large ESs for pre- to post-change (Cooper & Xu, 2021). |

| | | | | |
|---|---|---|--|--|
| The Values Wheel (VW; O'Connor et al., 2019) | The Values Wheel was developed for Acceptance and Commitment Therapy. It is administered by researchers and practitioners as a single interview-based instrument to assess the degree to which clients’ | The VW has been used with adult clinical populations. It is available in English. Available at: https://osf.io/8b6kr/ | No evidence of the clinical utility of the VW could be identified. | Convergent validity: in comparison to a range of standardised instruments relating to depression, anxiety, stress, and quality of life (AAQ-II; MHC-SF; DASS; SWLS; MCSD-SF; VQ; VLQ), significant Pearson’s correlations ranged from low to |
| | | | | (Continues) |

TABLE 2 (Continued)

| Name (key reference[s], website) | Brief description (inc. population, item generation, item rating, assessment, scoring). Variations. | Usage to date (who by, included in monitoring systems) | Clinical utility/face validity (acceptability, clinical usefulness, feasibility) | Psychometric evidence (reliability [internal, test-retest], convergent validity, divergent validity, sensitivity to change) |
|----------------------------------|--|--|--|--|
| | actions have been guided by their idiographic and weighted values over the previous week. The client writes a discrete personal value on the lip of each coloured disk. They then rate the extent to which their behaviour has been guided by each value over the past week by placing a mark on the target of each respective coloured disk to represent this. Finally, the client adjusts the area of the circle covered by each coloured disk to reflect the relative importance of each value. | | | moderate. Scores on the VW were positively related to outcomes on measures of well-being and life satisfaction, though no significant negative correlations were found for anxiety, stress, or depression. Discriminant validity: A significant negative correlation was seen in the relationship between VW and years of education. Convergent validity: Convergence with alternative measures of values was demonstrated. Incremental validity: VW accounted for unique proportions of variance in well-being, mental health and life satisfaction (Barrett et al., 2020). |

Abbreviation: I-PROM, Idiographic patient-reported outcome measures.

disagreements over the goals or tasks of the therapeutic work, providing an opportunity for repair (Bradley et al., 1999). Research also indicates that adult patients value the freedom given by I-PROMs to include problems/items of their own choosing (Alves et al., 2016).

Within the field of palliative healthcare, I-PROMs have been perceived as supporting communication by enabling patients to tell their own story in their own words (Greenhalgh et al., 2018). I-PROMs, in mental health treatments too, generate unique information about the patient's narrative, and therefore have been used by clinicians to complement diagnoses and for treatment planning (Antunes et al., 2020; Sales et al., 2007). In family therapy, for instance, the creation of individual problem questionnaires (PQs) for each family member can be used to "hold" multiple—and, at times, conflicting—views and needs of different family members (Sales et al., 2019). I-PROMs also provide an evidence-based, structured strategy to collect qualitative and personalized information about adult and youth patients that can be used for systematic evaluation purposes (unlike more informal methods, such as clinical notes), to support supervision and the preparation of case studies using sophisticated designs, such as the hermeneutic single-case efficacy design (Carvalho et al., 2008; Elliott et al., 2016).

I-PROMs allow case-specific feedback on clinical progress through the tracking of different areas of importance to the patient, both adult and youth. This is, again, consistent with the clinical reality of psychotherapeutic work, as patients may have the same score from the same nomothetic PROM, but be working towards different areas on I-PROMs based on their personal goals and difficulties (Ashworth et al., 2007; Edbrooke-Childs et al., 2015; Sales et al., 2018). Since I-PROMs allow monitoring progress on case-specific issues that are not captured by nomothetic tools (Ashworth et al., 2007; Sales et al., 2018) they introduce new pieces of information that are valued and used by therapists (Antunes et al., 2020; Barkham, 2016). For instance, an analysis of how therapists working with adults used PQ feedback showed that, instead of relying on overall change scores, they examined change in specific problem areas. This was coupled with using qualitative information to identify themes to explore in sessions, and tailoring intervention to meet the needs of each family member (Sales et al., 2018).

I-PROMs can be used by psychotherapists as part of a wider monitoring system to support session-to-session clinical decision making with both adult and youth patients. For instance, the Individualized Patient Progress System (IPPS; Sales et al., 2014)—co-designed by researchers and practitioners to monitor progress of individuals, families, and groups—integrates the PQ with the nomothetic CORE-Net system (Barkham et al., 2015). Similarly, the Marriage and Family Therapy Practice Research Network developed an internet-based assessment portal in which each family member was asked about their top three presenting problems, which were then scored for intensity at each session (Johnson et al., 2017).

Preliminary evidence from controlled studies suggests that the use of I-PROMs may be associated with improved outcomes in adults (McMurran et al., 2013; Mintz & Kiesler, 1982; Smith, 1994). For instance, McMurran et al. (2013) found that clients randomized to use of the Personal Concerns Inventory I-PROM after initial assessment, alongside treatment as usual, had a median session attendance of 88.3% over 12 weeks, compared to 66.7% attendance over the same period for clients receiving TAU only. In the same study, mean treatment engagement scores—as measured using the Treatment Engagement Rating scale (Drieschner & Boomsma, 2008)—were higher in the PCI group compared to those receiving TAU only (6.64 and 2.94, respectively). The use of goal-based I-PROMs is also reinforced by an abundance of evidence on the beneficial effects of personal goal setting in adults (Epton et al., 2017) and tracking of goal progress (Harkin et al., 2016).

1.1.2 | Patient-centeredness

Closely linked to clinical utility, the patient-centeredness of I-PROMs can serve to promote an individual rather than normative identity, for both adults and youth (Smith, 1994). This is important when considering the personal and fluctuating nature of personal change or mental health recovery (Barber et al., 2017; Corrigan et al., 1999; Jacobson & Greenley, 2001; H. Law et al., 2020; Onken et al., 2007). Research in the healthcare field indicates that

standardized PROMs could be perceived by clinicians as trivializing patients' emotions and leading to a mechanized process of "question bombardment" (Greenhalgh et al., 2018). In addition, thematic analysis of I-PROM free-text items indicates that more than 50% of I-PROM contents are not featured in the pre-set items of their nomothetic PROMs counterparts, creating considerable nonoverlap between these two modalities of outcome measurement (Alves et al., 2020; Ashworth et al., 2007; Sales et al., 2018). This may be why, when measuring change on items that are of key importance to the patient, I-PROMs change scores have consistently been found to exceed nomothetic PROMs change scores for both adults and youth (Ashworth et al., 2005; Edbrooke-Childs et al., 2015; Elliott et al., 2016; Jacob et al., 2021; Karpenko & Owens, 2013; Krause et al., 2020).

A recent realist synthesis of literature on PROMs to support clinician–patient communication in the healthcare field found that, "Those studies that directly compared individualised and standardised PROMs found that patients felt the former had greater validity and were less distressing" (Greenhalgh et al., 2018, p. 20; Neudert et al., 2001). By centering ROM on the perspective of the adult or youth patient, patients may feel empowered with an enhanced sense of agency. I-PROMs have the capacity to strengthen the patient's voice and their sense of building responsibility for their own care (Antunes et al., 2020; Tollefsen, Neumer, et al., 2020). The use of I-PROMs also ensures that the voice of adult or youth patients is considered seriously by psychological therapists and services (Department of Health, 2012; Sales & Alves, 2016; Sales, 2017). In this sense, I-PROMs do not limit patients based on their value-systems, culture, or conceptualization of treatment success (Jacob et al., 2015, 2017; Kiresuk et al., 1994). This may be particularly important as there are ongoing questions about the extent to which standardized measures are inclusive of patients from minoritized groups, especially if the measures have not been developed or adapted to such groups.

1.2 | Problems and challenges of I-PROMs

The use of I-PROMs in ROM raises several important theoretical and methodological challenges.

1.2.1 | Item generation

Some patients may indicate that they are not able to accurately identify their problems, or goals, at intake (Alves et al., 2016; Antunes et al., 2020; Di Malta et al., 2019). This may be due to lack of self-awareness or because talking about them is too stressful. Depressed patients, or patients facing severe psychiatric conditions, may experience interference with their ability to formulate goals/problems, and patients may only be able to do so once a degree of symptomatic remission has been achieved (Guerra et al., 2018). Further, it has been argued that the setting of problems or goals on I-PROMs is susceptible to "gaming", whereby patients may choose easily attainable areas to focus on (Bevan & Hood, 2006; D. Law & Jacob, 2015; Wolpert et al., 2015). Equally, practitioners or services may orientate patients towards more achievable goals, or more resolvable problems, to "evidence" greater change. With I-PROMs, item content and appropriateness are not controlled a priori, yet it will affect the outcomes.

Patients (or therapists) may also focus on topics that are not directly relevant to therapy. For instance, a patient may identify as a problem "My manager is too controlling", or "The world is becoming over-populated", which their treatment may not be designed to address. These items may be important because they inform the therapist about issues that affect the patient's quality of life, but it is not clear that they should be used for measuring therapeutic change. To deal with the quality of free-text items, Elliott (2012) developed a quality item rating system for the PQ that can be used for other I-PROMs, where an item is classified as "well-formed" if it describes a specific personal difficulty that is reasonably a focus for psychological therapies, and does not address general societal issues (such as over-population). However, further research is required to develop a systematic item–quality procedure.

Another issue is that patients' principal problems, or goals, may change over the course of therapy. This means that patients may be rating, or working to address, problems or goals that are no longer relevant, or may have been resolved or achieved. Some I-PROMs allow for modifications. For instance, the Goals Form (Cooper & Xu, 2021) and the PQ (Elliott et al., 2016) have procedures for deleting, and adding, goals or problems. PSYCHLOPS also allows adding a new problem in the session-to-session form (Czachowski et al., 2011). Patient-generated process measures (Sales & Alves, 2016), such as the Helpful Aspects of Therapy questionnaire (Llewelyn, 1988) can prove useful in developing I-PROMs that can assess micro, session-level outcomes as well as improvements in the primary presenting issue(s) (McLeod, 2017). However, this adds considerable complexity to the assessment of change over time, particularly in comparison with nomothetic PROMs.

1.2.2 | Measurement model

Considering their idiosyncrasy and heterogeneity, a fundamental issue is whether I-PROM items are indeed indicators of a unified construct or whether should they be treated separately as observed individual variables of client change. In the latter case, it is questionable whether item scores can be aggregated to compute a conceptually coherent global score. This, then, makes it difficult for any psychometric approach to validity or reliability based on either inter-individual, intra-individual, and inter-item variability.

If we assume that I-PROM is measuring a unified construct, a new question arises: what is the psychological construct being measured by I-PROMs? Given that items are very likely different for each person, does the I-PROM questionnaire measure a specific construct for each patient? Or is it possible to argue that it measures the same construct across people? If the former hypothesis is assumed, psychometric analysis based on inter-individual variation is not appropriate, and only intra-individual techniques would be acceptable.

A third related question addresses the measurement model underlying I-PROMs. A measurement model specifies the nature of the association between the obtained observations (indicators) and the theoretical latent construct being measured. Bollen and Diamantopoulos (2017) state that it is "crucial to correctly specify the type of indicators being used for measurement, as this determines the instrument metric properties, as well as its appropriate applications" (p. 3). We will briefly present the two principal measurement models in use for psychological measures, the *reflective* and *formative* models, and will proceed by explaining why, from our perspective, the I-PROMs measurement model is still immature, hindering their psychometric evaluation.

In the *reflective* model, the observed indicators (items) reflect—and are an effect of—the hypothesized latent construct (Bollen & Bauldry, 2011; Diamantopoulos & Winklhofer, 2001; Edwards & Bagozzi, 2000; Jarvis et al., 2003). This is the measurement model underlying classical test theory (e.g., Nunnally & Bernstein, 1994) and established nomothetic PROMs where, for instance, scores on an item such as "How often have you been bothered by feeling down, depressed, or hopeless?" are considered reflective of an underlying, latent condition or construct (e.g., depression) (Kroenke et al., 2001). Change in the items reflects change in the construct. By contrast, in the *formative* measurement model, the observed indicators form—in the sense that they cause—the hypothesized construct: for instance, when assessing the construct life stress by inquiring about an illness in the family. Here, change in the item (e.g., family member illness) causes change in the construct life stress.

Adopting one measurement model over the other requires different approaches in psychometrics. In particular, with the formative model, different causal indicators are not necessarily expected to correlate with each other. Hence, techniques relying on shared variance (e.g., factor analysis), or internal consistency based on inter-item correlations (e.g., coefficient alpha), may not be appropriate for the evaluation of formative-based measures (Diamantopoulos et al., 2008). As I-PROM items are generated by patients, they may be both reflective of an underlying latent construct, or formative of it. Indeed, most probably, I-PROMs would combine both reflective and formative problems/goals. It is difficult to classify I-PROMs in terms of their measurement model, and therefore its psychometric evaluation can be problematic.

In our view, these three issues regarding the measurement model of I-PROMs constitute the major methodological challenge to the idiographic assessment of outcomes in psychological treatment. There have been methodological proposals to deal with these issues (e.g., Elliott et al., 2016; Gaasterland et al., 2019; Urach et al., 2018). However, we do not have a consensus answer to these questions yet, as the psychometric science of I-PROMs continues to emerge. Our recommendation is that any psychometric analysis of I-PROMs explicates and justifies the measurement model assumed, along with the corresponding data analysis plan (see Sales et al., 2021). Further, we propose that analyses should preferably use methods that are applicable to both reflective and formative measurement models. Accordingly, we display in Table 3 key methods used to evaluate the reliability and validity of measures, arranged in descending order according to their applicability to the two models. We also make specific recommendations in the next sections.

1.3 | Establishing reliability

The prevalent method for reliability assessment used on reflective models—internal consistency through coefficients such as Cronbach's alpha, once the factor structure of the measure has been established—should not be blindly adopted with I-PROMs. Internal consistency only applies to the reflective measurement model for unidimensional (sub)scales (COSMIN guidelines; Prinsen et al., 2018). In the formative model, the correlation of indicators can be high, or low, or null (Coltman et al., 2008; Edwards, 2011), so low intercorrelation of indicators does not mean low reliability. I-PROM items can even be considered a set of observed variables which do not have conceptual unity (Bollen & Bauldry, 2011). Therefore, Cronbach's alpha should be used with I-PROMs only when the reflective measurement model can be explicitly assumed (e.g., PSYCHLOPS in Sales et al., 2021).

Since item scores should be stable across time in both the reflective and formative models (Diamantopoulos et al., 2008), test–retest reliability is still applicable to I-PROMs that assume that items are indicators (either reflective or formative) of a psychological construct. However, test–retest computation requires between-individuals variation, so it can only be applied to I-PROMs that measure the same construct across patients.

There are alternatives to examine reliability of I-PROMs at the intra-individual level, when there is data available over time (e.g., session-to-session intermediate outcomes). Cronbach's alpha can be computed based on item covariances estimated at the intra-individual level; session-to-session autocorrelations (lag-1) can also be used to assess temporal consistency within patients (Elliott et al., 2016). The reliability across clients, in terms of mean reliabilities and ranges, can then be reported at the population level (e.g., in PQ, Elliott et al., 2016). This intra-individual approach, however, is not without limitations when applied to isolated items since it is not possible to distinguish if an observed score variation over time in the target item is due to measurement error (i.e., informs about the reliability of the tool) or to real therapeutic change (i.e., informs about the tool's sensitivity to change).

Finally, I-PROMs do not include only scoring but also patient-indicated qualitative data, which calls for the evaluation of item's content reliability. This is particularly important in self-report I-PROMs, such as PSYCHLOPS, where patients may provide an incomplete/ambiguous description of the problem (e.g., “my mother”). Here we recommend that therapists clarify the content and together with the patient reformulate the item in a way that clarifies its meaning. Furthermore, quality and reporting standards for qualitative research (Elliott, 1991; Levitt et al., 2018) can be used in. Indeed, if I-PROM items are to be categorized based on content, it is particularly important to ensure categorization is transparent and reproducible; for example, being conducted by, and consistent across, more than one researcher. Categories should be reviewed by clinicians and patients to ensure they are meaningful and appropriate to different interpretations of what a particular problem/goal might mean. Moreover, future research should examine the distinctiveness of I-PROM item categories and whether different demographic and clinical characteristics are associated with setting different types of item categories, and whether certain item categories show higher or lower levels of progress. This would further help to ensure the reliability of I-PROMs and the triangulation of the qualitative and quantitative data they uniquely afford.

TABLE 3 Methods to analyze the psychometric properties of I-PROMs

| Psychometric Properties | Content | Statistics | Applicability | Level | Methods | Limitations |
|-------------------------|-------------------------|--|---|------------------|--|--|
| Reliability | Test-retest reliability | Correlations of either measures or items at two different time points | Both reflective model and formative model | Intra-individual | Multilevel modeling | Requires multi-items as indicators of the same construct |
| | | | | Interindividual | Bivariate correlation | Requires the assumption that scores/items are measuring the same construct between individuals |
| | Internal consistency | Cronbach's α | Reflective model. Not appropriate to formative model: high positive item intercorrelation is unnecessary | Intra-individual | Cronbach's α for each patient (mean and 95% CI of α s) | Requires multi-items as indicators of the same construct Unsuitable for patients with few sessions. Therefore, not all patients could be included in analysis |
| Validity | Convergent validity | Correlations between either measures or items and other nomothetic outcome measures. | Both reflective model and formative model. It is expected the latent variable completely mediates the effects of its indicators on other nomothetic outcome variables in formative model. | Interindividual | | Not appropriate and meaningful for I-PROMs, unless a reflective model is justifiably assumed, which considers that items are measuring the same construct between individuals. |
| | | | | Intra-individual | Multilevel modeling | Requires the assumption that scores/items are measuring the same construct between individuals. |
| | Discriminant validity | Correlations between either measures or items and other | Both reflective model and formative model. | Intra-individual | Multilevel modeling | Requires the assumption that scores/items are measuring |
| | | | | Interindividual | Bivariate correlation/ | |

(Continues)

TABLE 3 (Continued)

| Psychometric Properties | Content | Statistics | Applicability | Level | Methods | Limitations |
|-------------------------|---------|--|---|------------------|------------------------------------|--|
| Construct validity | | theoretically irrelevant variables. | | | Structure equation modeling | the same construct between individuals |
| | | Loadings of the indicators to the construct. | Both reflective model and formative model, not all the loadings are expected high in formative model. | Intra-individual | P-Factor analysis for each patient | Unsuitable for patients with few sessions. Therefore, not all patients could be included in analysis |
| | | | | Interindividual | | Not appropriate and meaningful for I-PROMs, unless a reflective model is justifiably assumed, which considers that items are measuring the same construct between individuals. |

Abbreviation: I-PROM, Idiographic patient-reported outcome measures.

1.4 | Establishing validity

For construct validity in nomothetic reflective measures, common factors are predominantly studied via factor analysis. For I-PROMs, P-techniques for factor analysis (Molenaar & Nesselroade, 2009) applied to item covariances at the intra-individual level (across sessions) can be used to extract or confirm latent dimensions. However, such an approach is still based on the assumption of highly intercorrelated items, where the percentage of shared variance and communalities are emphasized. In formative models, by contrast, high covariance among items is not required, and construct validity is assessed by the correlation between indicators and construct—that is, the contribution of the individual indicators to the constructs (Diamantopoulos et al., 2008). Both approaches assume that I-PROM is measuring a latent construct.

Convergent and divergent validity of I-PROMs may be explored at the intra-individual level (for instance, bivariate correlation between I-PROMs scores, or even single/isolated I-PROM items, with an external criterion across sessions). However, since I-PROMs may be considered different for each patient, convergent and divergent validity approaches based on inter-individual differences should be used carefully and exclusively when a justifiable rationale exists.

According to the COSMIN guidelines, the most important measurement property of a psychological scale is content validity (Prinsen et al., 2018), which is theoretically ensured by the fact that I-PROM items are indicated by the patient.

1.4.1 | Service evaluation and aggregated scores

A key question is the extent to which I-PROMs can be used to compare outcomes of patients at the service level. This is a critical consideration for the take-up of such measures by commissioners. We argue that I-PROM data can be used at both the individual and service level (Edbrooke-Childs et al., 2015). Suppose a service brought about reductions in distress but, overall, no one reached their personal goals or solved their personal problems. Is that good? To the extent that we hold open the question of what therapy is for, we propose that I-PROM scores are needed as population-level indicators of service quality.

Service-level data may be analyzed following the same procedure as are used for PROMs—i.e., computing the total score and calculating the pre–post clinically significant change (Elliott et al., 2016; Sales et al., 2021). Alternative methods have also been developed to accommodate the specific nature of I-PROM data. Currently, for instance, in Child and Adolescent Mental Health Services in England, NHS England and NHS Improvement are trialing an outcomes metric which incorporates change on at least one goal on the GBO tool, utilizing the principles of the reliable change index (Jacob et al., 2021; Jacobson & Truax, 1991). This outcome metric is in its “shadow year” and its findings will be watched with interest. In analysis of service outcomes based on this principle, the combining of data from an I-PROM with data from nomothetic measures gave higher levels of reliable improvement compared to nomothetic measures alone. This suggests that important improvements may be missed without the use of I-PROMs data at the service level (Jacob et al., 2021).

Further data analysis procedures have been developed to deal with the challenges of inter-individual comparison of I-PROMs. For instance, the Metric-Frequency Similarity Index (Sales & Wakker, 2009; Sales et al., 2015) was developed to compare PQ questionnaires, in a way that considers the content and the intensity of all free-text items. Its algorithm was implemented in feedback systems to compare members of a therapeutic group and family members regarding PQ (Sales et al., 2014). Another example is the use of Group Iterative Multiple Model Estimation (GIMME) to integrate both I-PROM and nomothetic measures (Beltz & Gates, 2017; Beltz et al., 2016; Gates et al., 2019). This method is based on modelling the relationships between a set of variables for individuals over time, creating individual-level networks that include some group-level relations, therefore generating associations that are shared by all individuals in a sample. GIMME generates a graph for each participant that can be

conceptualized as a person-specific network or connectivity map. Further, this method accommodates group-level information in the individual level networks multilevel structural equation techniques (Beltz et al., 2016; Wright et al., 2015). Similarly, analysis of patient-indicated goals by its classification into pre-determined, data driven themes has been suggested as useful for identifying training needs and for service planning (Jacob et al., 2021).

1.5 | Final comments

Given the epistemological differences on which I-PROMs and nomothetic PROMs are based, it is likely that these different types of measurement will appeal to—and sit within—differing practitioners' and researchers' epistemological positionings. However, it should be recognized that science is not purely nomothetic or idiographic. To maximize the ability to understand change in terms of individuals' lived experiences, both may be necessary (Allport, 1961; Ashworth et al., 2019; Edbrooke-Childs et al., 2015; Green, 2016; Sales et al., 2007; Wolpert et al., 2015). Hence, while theoretical arguments may accentuate the epistemological differences, these may be less evident in practice (Ashworth et al., 2019).

Current knowledge about the metric characteristics and quality standards of I-PROMS is, however, insufficient. Further research is required, for instance, on how to generate items so that they are specific and realistic in the context of psychological therapy (Cooper, 2019). Similarly, there is the question of how to accommodate changes to the content of I-PROMs over the course of treatment (e.g., setting new goals/problems, or deleting problems), and how this might be incorporated into any aggregate score. Further work is also needed on the appropriateness of applying statistical methods to analyze the psychometric properties of I-PROMs, based on its congruence with the measurement model of each I-PROM tool. In our view, it is crucial that all the methodological and analytic solutions are reported in detail, so that they can be contrasted and evaluated in a future systematic review.

Despite these ongoing questions, the importance of including I-PROMs in ROM for psychological therapies is clear. Tools designed to capture ongoing therapeutic change must show acceptability, feasibility, and face validity, as seen by key stakeholder groups: in this case, patients and therapists. In their realist synthesis on PROMs and clinician–patient communication in the healthcare field, Greenhalgh et al. (2018) found:

clinicians across a range of clinical settings found using a standardised PROM during initial assessments could constrain, rather than support communication and interfered with the process of managing relationships with patients, while individualised PROMs supported this dialogue. (p. 23)

I-PROMs' unique strengths of clinical utility and person-centredness (with a content validity inherently high from the individual patient's perspective) puts them as equivalent and complementary to nomothetic PROMs for monitoring change in routine clinical settings. Our recommendation is that I-PROMs are combined with nomothetic PROMs—balancing the strengths and limitations of each type, alone, following an integrated measurement approach called personalized ROM (Sales & Alves, 2012).

Attempts to resolve epistemological, empirical, and practical challenges should not come at the expense of compromising the value of these tools: ensuring the individual patient's voice is central to the psychotherapeutic process and assessment of its outcomes even if that voice comes in many different tones.

ETHICS APPROVAL STATEMENT

As a position paper reporting only on secondary data, no ethical approval was required for this paper.

DATA AVAILABILITY STATEMENT

This manuscript did not use empirical data.

ORCID

Célia M. D. Sales  <http://orcid.org/0000-0002-2586-8469>

Mark Ashworth  <http://orcid.org/0000-0001-6514-9904>

Salma Ayis  <https://orcid.org/0000-0002-0449-869X>

Michael Barkham  <https://orcid.org/0000-0003-1687-6376>

Julian Edbrooke-Childs  <https://orcid.org/0000-0003-0401-4058>

Luís Faisca  <https://orcid.org/0000-0003-4859-8817>

Jenna Jacob  <http://orcid.org/0000-0003-1006-1547>

Mick Cooper  <http://orcid.org/0000-0003-1492-2260>

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1002/jclp.23319>

REFERENCES

- Allport, G. W. (1961). *Personality: A psychological interpretation*. Holt, Rinehart & Winston.
- Alves, P. C. G., Sales, C. M. D., & Ashworth, M. (2016). "It is not just about the alcohol": Service users' views about individualised and standardised clinical assessment in a therapeutic community for alcohol dependence. *Substance Abuse Treatment, Prevention, and Policy*, 11(1), 25. <https://doi.org/10.1186/s13011-016-0070-5>
- Alves, P. C. G., Sales, C. M. D., Ashworth, M., & Faisca, L. (2020). "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*, 18(4), 861–874. <https://doi.org/10.1007/s11469-018-9985-6>
- Antunes, R. P., Sales, C. M. D., & Elliott, R. (2020). The clinical utility of the Personal Questionnaire (PQ): A mixed methods study. *Counselling Psychology Quarterly*, 33(1), 25–45. <https://doi.org/10.1080/09515070.2018.1439451>
- Ashworth, M., Evans, C., & Clement, S. (2009). Measuring psychological outcomes after cognitive behaviour therapy in primary care: A comparison between a new patient-generated measure "PSYCHLOPS" (Psychological Outcome Profiles) and "HADS" (Hospital Anxiety and Depression Scale). *Journal of Mental Health*, 18(2), 169–177. <https://doi.org/10.1080/09638230701879144>
- 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., Robinson, S., Evans, C., Shepherd, M., Conolly, A., & Rowlands, G. (2007). What does an idiographic measure (PSYCHLOPS) tell us about the spectrum of psychological issues and scores on a nomothetic measure (CORE-OM)? *Primary Care & Community Psychiatry*, 12(1), 7–16. <https://doi.org/10.1080/17468840701560805>
- Ashworth, M., Robinson, S., Godfrey, E., Shepherd, M., Evans, C., Seed, P., Parmentier, H., & Tylee, A. (2005). Measuring mental health outcomes in primary care: The psychometric properties of a new patient-generated outcome measure, "PSYCHLOPS" ('psychological outcome profiles'). *Primary Care Mental Health*, 3(4), 261–270.
- 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.
- Barber, J. M., Parsons, H., Wilson, C. A., & Cook, C. C. H. (2017). Measuring mental health in the clinical setting: What is important to service users? The Mini-Service user Recovery Evaluation scale (Mini-SeRvE). *Journal of Mental Health*, 26(6), 530–537. <https://doi.org/10.1080/09638237.2017.1340624>
- Barkham, M. (2016). Patient-centered assessment in psychotherapy: Toward a greater bandwidth of evidence. *Clinical Psychology: Science and Practice*, 23(3), 284–287. <https://doi.org/10.1111/cpsp.12163>
- Barkham, M., Hardy, G. E., & Mellor-Clark, J. (2010). Developing and delivering practice-based evidence, *A guide for the psychological therapies*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470687994>
- Barkham, M., Leach, C., Lucock, M., Evans, C., Margison, F., Mellor-Clark, J., Benson, L., Connell, J., Audin, K., & 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>
- Barkham, M., Mellor-Clark, J., & Stiles, W. B. (2015). A CORE approach to progress monitoring and feedback: Enhancing evidence and improving practice. *Psychotherapy*, 52(4), 402–411. <https://doi.org/10.1037/pst0000030>
- Barrett, K., O'Connor, M., & McHugh, L. (2020). Investigating the psychometric properties of the Values Wheel with a clinical cohort: A preliminary validation study. *Journal of Contextual Behavioral Science*, 15, 210–218. <https://doi.org/10.1016/j.jcbs.2019.12.009>

- Baumann, N. (2011). Volitional and emotional correlates of the motivational. In W. M. Cox, & E. Klinger (Eds.), *Structure Questionnaire: Further Evidence for Construct Validity* (2nd ed, pp. 233–252). Handbook of Motivational Counseling. <https://doi.org/10.1002/9780470979952.ch9>
- Beltz, A. M., & Gates, K. M. (2017). Network mapping with GIMME. *Multivariate Behavioral Research*, 52(6), 789–804. <https://doi.org/10.1080/00273171.2017.1373014>
- Beltz, A. M., Wright, A. G. C., Sprague, B. N., & Molenaar, P. C. M. (2016). Bridging the nomothetic and idiographic approaches to the analysis of clinical data. *Assessment*, 23(4), 447–458. <https://doi.org/10.1177/1073191116648209>
- Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration*, 84(3), 517–538. <https://doi.org/10.1111/j.1467-9299.2006.00600.x>
- Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods*, 16(3), 265–284. <https://doi.org/10.1037/a0024448>
- 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>
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice*, 16(3), 252–260. <https://doi.org/10.1037/h0085885>
- Bradley, E. H., Bogardus, S. T. J., Tinetti, M. E., & Inouye, S. K. (1999). Goal-setting in clinical medicine. *Social Science & Medicine*, 49(2), 267–278. [https://doi.org/10.1016/s0277-9536\(99\)00107-0](https://doi.org/10.1016/s0277-9536(99)00107-0)
- Bromley, C., & Westwood, S. (2013). Young people's participation: Views from young people on using goals. *Clinical and Family Psychology Review*, 1, 41–60.
- Cairns, A. J., Kavanagh, D. J., Dark, F., & McPhail, S. M. (2019). Goal setting improves retention in youth mental health: A cross-sectional analysis. *Child and Adolescent Psychiatry And Mental Health*, 13(1), 31–39. <https://doi.org/10.1186/s13034-019-0288-x>
- Cardillo, J. E., & Smith, A. (1994). Reliability of goal attainment scores, In T. J. Kiresuk, A. Smith, & J. E. Cardillo (Eds.), *Goal attainment scaling: Applications, theory, and measurement* (pp. 213–241). Lawrence Erlbaum Associates.
- Carey, K. B., Leontieva, L., Dimmock, J., Maisto, S. A., & Batki, S. L. (2007). Adapting motivational interventions for comorbid schizophrenia and alcohol use disorders. *Clinical Psychology: Science and Practice*, 14(1), 39–57. <https://doi.org/10.1111/j.1468-2850.2007.00061.x>
- Carvalho, M. J., Faustino, I., Nascimento, A., & Sales, C. M. D. (2008). Understanding Pamina's recovery: An application of the hermeneutic single-case efficacy design. *Counselling and Psychotherapy Research*, 8(3), 166–173. <https://doi.org/10.1080/14733140802211002>
- Clark, D. M. (2018). Realizing the mass public benefit of evidence-based psychological therapies: The IAPT program. *Annual Review of Clinical Psychology*, 14(1), 159–183. <https://doi.org/10.1146/annurev-clinpsy-050817-084833>
- Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, 61(12), 1250–1262. <https://doi.org/10.1016/j.jbusres.2008.01.013>
- Cooper, M. (2019). Integrating counselling and psychotherapy: Directionality, synergy, and social change. Sage.
- Cooper, M., & McLeod, J. (2011). Person-centered therapy: A pluralistic perspective. *Person-Centered & Experiential Psychotherapies*, 10(3), 210–223. <https://doi.org/10.1080/14779757.2011.599517>
- Cooper M., & Xu, D. (2021). The Goals Form: Reliability, validity, and clinical utility of an idiographic goal-focused measure for routine outcome monitoring in psychotherapy. *Journal of Clinical Psychology*.
- Cooper, M., Wild, C., van Rijn, B., Ward, T., McLeod, J., Cassar, S., Antoniou, P., Michael, C., Michalitsi, M., & Sreenath, S. (2015). Pluralistic therapy for depression: Acceptability, outcomes and helpful aspects in a multisite study. *Counselling Psychology Review*, 30(1), 6–20.
- Corrigan, P. W., Giffort, D., Rashid, F., Leary, M., & Okeke, I. (1999). Recovery as a Psychological Construct. *Community Mental Health Journal*, 35(3), 231–239. <https://doi.org/10.1023/A:1018741302682>
- Cox, W. M., & Klinger, E. (2021). Assessing current concerns and goals idiographically: A review of the motivational structure questionnaire family of instruments. *Journal of Clinical Psychology*, 1–16. <https://doi.org/10.1002/jclp.23256>
- Czachowski, S., Seed, P., Schofield, P., & Ashworth, M. (2011). Measuring psychological change during cognitive behaviour therapy in primary care: A Polish study using "PSYCHLOPS" (Psychological Outcome Profiles). *PLoS One*, 6(12), e27378. <https://doi.org/10.1371/journal.pone.0027378>
- Department of Health. (2012). Liberating the NHS: No decision about me, without me. NHS. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/216980/Liberating-the-NHS-No-decision-about-me-without-me-Government-response.pdf
- Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, 61(12), 1203–1218. <https://doi.org/10.1016/j.jbusres.2008.01.009>

- Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38(2), 269–277. <https://doi.org/10.1509/jmkr.38.2.269.18845>
- Dowden, C., Chambers, N., Hunt, J., Richardson-Taylor, K., Hargrave, A., & Little, B. (2001). Personal project appraisal and depressive affect: A meta-analysis (Unpublished Manuscript), Carleton University, Ottawa, Canada.
- Drieschner, K. H., & Boomsma, A. (2008). The Treatment Motivation Scales for forensic outpatient treatment (TMS-F): Construction and psychometric evaluation. *Assessment*, 15(2), 224–241. <https://doi.org/10.1177/1073191107311650>
- 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>
- Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational Research Methods*, 14(2), 370–388. <https://doi.org/10.1177/1094428110378369>
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5(2), 155–174. <https://doi.org/10.1037/1082-989X.5.2.155>
- Elliott, R. (1991). Five dimensions of therapy process. *Psychotherapy Research*, 1(2), 92–103. <https://doi.org/10.1080/10503309112331335521>
- Elliott, R. (2012). *Quality rating system for PQ items*. Unpublished manuscript. Counselling Unit, University of Strathclyde.
- Elliott, R., Wagner, J., Sales, C. M. D., Alves, P. C. G., Rodgers, B., & Café, M. (2016). Psychometrics of the Personal Questionnaire: A client-generated outcome measure. *Psychological Assessment*, 28(3), 263–278. <https://doi.org/10.1037/pas0000174>
- Emmons, R. A. (1986). Personal strivings: An approach to personality and subjective well-being. *Journal of Personality and Social Psychology*, 51(5), 1058–1068. <https://doi.org/10.1037/0022-3514.51.5.1058>
- Epton, T., Currie, S., & Armitage, C. J. (2017). Unique effects of setting goals on behavior change: Systematic review and meta-analysis. *Journal of Consulting and Clinical Psychology*, 85(12), 1182–1198. <https://doi.org/10.1037/ccp0000260>
- Evans, C. (2012). Cautionary notes on power steering for psychotherapy. *Canadian Psychology-Psychologie Canadienne*, 53(2), 131–139. <https://doi.org/10.1037/a0027951>
- Evans, C., Ashworth, M., & Peters, M. (2010). Are problems prevalent and stable in non-clinical populations? Problems and test-retest stability of a patient-generated measure, PSYCHLOPS (Psychological Outcome Profiles), in a non-clinical student sample. *British Journal of Guidance & Counselling*, 38(4), 431–439. <https://doi.org/10.1080/03069885.2010.503701>
- Feltham, A., Martin, K., Walker, L., & Harris, L. (2018). Using goals in therapy: The perspective of people with lived experience. In M. Cooper, & D. Law (Eds.), *Working with goals in psychotherapy and counselling* (pp. 73–85). Oxford University Press.
- Gaasterland, C. M. W., van der Weide, M. C. J., Roes, K. C. B., & van der Lee, J. H. (2019). Goal attainment scaling as an outcome measure in rare disease trials: A conceptual proposal for validation. *BMC Medical Research Methodology*, 19(1), 227. <https://doi.org/10.1186/s12874-019-0866-x>
- Gates, K., Fisher, Z., & Bollen, K. (2019). Latent variable GIMME using model implied instrumental variables (MIIVs). *Psychological Methods*, 25(2), 227–242. <https://doi.org/10.1037/met0000229>
- 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. (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>
- Greenhalgh, J., Gooding, K., Gibbons, E., Dalkin, S., Wright, J., Valderas, J., & Black, N. (2018). How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. *Journal of Patient-Reported Outcomes*, 2(1), 42. <https://doi.org/10.1186/s41687-018-0061-6>
- Grosse, M., & Grawe, K. (2002). Bern Inventory of Treatment goals: Part 1. Development and first application of a taxonomy of treatment goal themes. *Psychotherapy Research*, 12(1), 79–99. <https://doi.org/10.1080/713869618>
- Guerra, D., Pereira, J., & Sales, C. M. D. (2018). Assessment of outcomes at Casa de Alba: Learning from the residents' views. In J. Pereira, J. Gonçalves, & V. Bizzari (Eds.), *The neurobiology. psychotherapy, pharmacology intervention triangle: The need for integration in 21st century mental health*, 205–219. Vernon Press.
- Harkin, B., Webb, T. L., Chang, B. P. I., Prestwich, A., Conner, M., Kellar, I., Benn, Y., & Sheeran, P. (2016). Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. *Psychological Bulletin*, 142(2), 198–229. <https://doi.org/10.1037/bul0000025>
- Jacob, J., Edbrooke-Childs, J., Costa da Silva, L., & Law, D. (2021). Notes from the youth mental health field: Using movement towards goals as a potential indicator of service change and quality improvement. *Journal of Clinical Psychology*. 1–14. <https://doi.org/10.1002/jclp.23195>

- Jacob, J., Edbrooke-Childs, J., Holley, S., Law, D., & Wolpert, M. (2015). 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. (2017). 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., De Francesco, D., Deighton, J., Law, D., Wolpert, M., & Edbrooke-Childs, J. (2017). 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>
- Jacobson, N. S., & Greenley, D. (2001). What is recovery? A conceptual model and explication. *Psychiatric Services*, 52(4), 482–485. <https://doi.org/10.1176/appi.ps.52.4.482>
- Jacobson, N. S., & Truax, P. (1991). 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//0022-006x.59.1.12>
- Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, 30(2), 199–218. <https://doi.org/10.1086/376806>
- Jensen-Doss, A., Smith, A. M., Becker-Haimes, E. M., Mora Ringle, V., Walsh, L. M., Nanda, M., Walsh, S. L., Maxwell, C. A., & Lyon, A. R. (2018). Individualized progress measures are more acceptable to clinicians than standardized measures: Results of a national survey. *Administration and Policy in Mental Health*, 45(3), 392–403. <https://doi.org/10.1007/s10488-017-0833-y>
- Johnson, L. N., Miller, R. B., Bradford, A. B., & Anderson, S. R. (2017). The Marriage and Family Therapy Practice Research Network (MFT-PRN): Creating a more perfect union between practice and research. *Journal of Marital and Family Therapy*, 43(4), 561–572. <https://doi.org/10.1111/jmft.12238>
- De Jong-Meyer, R. (2004). Systematic motivational analysis as part of a self-help technique aimed at personal goal attainment. In W. M. Cox, & E. Klinger (Eds.), *Handbook of motivational counseling* (pp. 259–282). John Wiley & Sons Ltd. <https://doi.org/10.1002/9780470713129>
- Karpenko, V., & Owens, J. S. (2013). Adolescent psychotherapy outcomes in community mental health: How do symptoms align with target complaints and perceived change? *Community Mental Health Journal*, 49(5), 540–552. <https://doi.org/10.1007/s10597-012-9515-0>
- Kashdan, T. B., Breen, W. E., & Julian, T. (2010). Everyday strivings in war veterans with Posttraumatic Stress Disorder: Suffering from a hyper-focus on avoidance and emotion regulation. *Behavior Therapy*, 41(3), 350–363. <https://doi.org/10.1016/j.beth.2009.09.003>
- 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>
- Kiresuk, T. J., Smith, C. C. J., & Cardillo, J. E. (1994). *Goal attainment scaling: Applications, theory, and measurement*. Lawrence Erlbaum.
- Klinger, E., & Cox, W. M. (2011). The Motivational Structure Questionnaire, Personal Concerns Inventory, and their variants: Psychometric properties. In W. M. Cox, & E. Klinge (Eds.), *Handbook of Motivational Counseling* (pp. 2nd ed. 205–232). Wiley. <https://doi.org/10.1002/9780470979952.ch8>
- Krause, K. R., Edbrooke-Childs, J., Bear, H. A., Calderon, A., & Wolpert, M. (2020). What treatment outcomes matter most? A Q-study of outcome priority profiles among youth with lived experience of depression. *MedRxiv*. <https://doi.org/10.1101/2020.10.12.20210468>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine*, 16(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Lambert, M., Whipple, J., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy*, 55(4), 520–537. <https://doi.org/10.1037/pst0000167>
- Law, D., & Jacob, J. (2015). Goals and goal based outcomes: Some useful information (3rd ed.). CAMHS Press.
- Law, H., Gee, B., Dehmahdi, N., Carney, R., Jackson, C., Wheeler, R., Carroll, B., Tully, S., & Clarke, T. (2020). What does recovery mean to young people with mental health difficulties? - "It's not this magical unspoken thing, it's just recovery". *Journal of Mental Health*, 29(4), 464–472. <https://doi.org/10.1080/09638237.2020.1739248>
- Levitt, H. M., Bamberg, M., Creswell, J. W., Frost, D. M., Josselson, R., & Suárez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 26–46. <https://doi.org/10.1037/amp0000151>
- 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>

- Little, B. R. (2011). Personal projects and motivational counseling: The quality of lives reconsidered. In W. M. Cox & E. Klinger (Eds.), *Handbook of Motivational Counseling* (2nd ed., pp. 73–87). Wiley. <https://doi.org/10.1002/9780470979952.ch3>
- Little, B. R., & Coulombe, S. (2015). Personal Projects. In J. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., pp. 757–765). Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.26100-X>
- Little, B. R., & Gee, T. L. (2007). Personal projects analysis. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (Vols. 1–3). SAGE Publications, Inc. <https://doi.org/10.4135/9781412952644.n343>
- Llewellyn, S. P. (1988). Psychological therapy as viewed by clients and therapists. *British Journal of Clinical Psychology*, 27(3), 223–237. <https://doi.org/10.1111/j.2044-8260.1988.tb00779.x>
- 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*, 26(3), e12281. <https://doi.org/10.1111/cpsp.12281>
- Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., Noble, R., & 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>
- Di Malta, G., 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>
- Man, F., Stuchliková, I., & Klinger, E. (1998). Motivational structure of alcoholic and nonalcoholic Czech men. *Psychological Reports*, 82(3_suppl), 1091–1106. <https://doi.org/10.2466/pr0.1998.82.3c.1091>
- McGaghie, W. C., & Menges, R. J. (1975). Assessing self-directed learning. *Teaching of Psychology*, 2(2), 56–59. https://doi.org/10.1207/s15328023top0202_2
- McLeod, J. (2017). Qualitative methods for routine outcome measurement, In T. Rousmaniere, R. K. Goodyear, S. D. Miller, & B. E. Wampold (Eds.), *The cycle of excellence: Using deliberate practice to improve supervision and training* (pp. 99–122). Wiley.
- McMurrin, M., Cox, W. M., Whitham, D., & Hedges, L. (2013). The addition of a goal-based motivational interview to treatment as usual to enhance engagement and reduce dropouts in a personality disorder treatment service: Results of a feasibility study for a randomized controlled trial. *Trials*, 14(1), 98. <https://doi.org/10.1186/1745-6215-14-50>
- Mintz, J., & Kiesler, D. J. (1982). Individualized measures of psychotherapy outcome. In P. C. Kendall, & J. N. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 491–534). John Wiley.
- Molenaar, P., & Nesselroade, J. (2009). The recoverability of P-technique factor analysis. *Multivariate Behavioral Research*, 44, 130–141. <https://doi.org/10.1080/00273170802620204>
- Morgan, M. A. J., Coates, M. J., Dunbar, J. A., Reddy, P., Schlicht, K., & Fuller, J. (2013). The TrueBlue model of collaborative care using practice nurses as case managers for depression alongside diabetes or heart disease: A randomised trial. *BMJ Open*, 3(1), e002171. <https://doi.org/10.1136/bmjopen-2012-002171>
- Neudert, C., Wasner, M., & Borasio, G. D. (2001). Patients' assessment of quality of life instruments: A randomised study of SIP, SF-36 and SEIQoL-DW in patients with amyotrophic lateral sclerosis. *Journal of the Neurological Sciences*, 191(1–2), 103–109. [https://doi.org/10.1016/S0022-510X\(01\)00612-8](https://doi.org/10.1016/S0022-510X(01)00612-8)
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- O'Connor, M., Tennyson, A., Timmons, M., & McHugh, L. (2019). The development and preliminary psychometric properties of the Values Wheel. *Journal of Contextual Behavioral Science*, 12, 39–46. <https://doi.org/10.1016/j.jcbs.2019.01.005>
- Onken, S. J., Craig, C. M., Ridgway, P., Ralph, R. O., & Cook, J. A. (2007). An analysis of the definitions and elements of recovery: A review of the literature. *Psychiatric Rehabilitation Journal*, 31(1), 9–22. <https://doi.org/10.2975/31.1.2007.9.22>
- van Orden, M., Hoffman, T., Haffmans, J., Spinhoven, P., & Hoencamp, E. (2009). Collaborative mental health care versus care as usual in a primary care setting: A Randomized Controlled Trial. *Psychiatric Services*, 60(1), 74–79. <https://doi.org/10.1176/ps.2009.60.1.74>
- Piccirillo, M. L., & Rodebaugh, T. L. (2019). Foundations of idiographic methods in psychology and applications for psychotherapy. *Clinical Psychology Review*, 71, 90–100. <https://doi.org/10.1016/j.cpr.2019.01.002>
- Prinsen, C. A. C., Mookink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1147–1157. <https://doi.org/10.1007/s11136-018-1798-3>
- Rajkarnikar, K. (2009). Assessment of social anxiety clients' symptoms: What do they really represent? University of Strathclyde/Glasgow Caledonian University.
- Sales, C. M. D. (2017). Seeing the person in the patient: Making the case for Individualized PROMs in mental health care. *Current Psychiatry Reviews*, 13(3), 184–187. <https://doi.org/10.2174/1573400513666170505111736>
- Sales, C. M. D., & Alves, P. C. G. (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. D., & Alves, P. C. G. (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. M. D., Alves, P. C. G., Evans, C., & Elliott, R. (2014). The individualised patient-progress system: A decade of international collaborative networking. *Counselling and Psychotherapy Research*, 14(3), 181–191. <https://doi.org/10.1080/14733145.2014.929417>
- Sales, C. M. D., Fáisca, L., Ashworth, M., & Ayis, S. (2021). The psychometric properties of PSYCHLOPS, an individualized patient-reported outcome measure of personal distress. *Journal of Clinical Psychology*. <https://doi.org/10.1002/jclp.23278>
- Sales, C. M. D., Ferreira, S., & Matos, P. M. (2019). How routine patient-centered monitoring relates to therapeutic gains in family therapy: A single-case study. *Journal of Marital and Family Therapy*, 45(4), 606–620. <https://doi.org/10.1111/jmft.12359>
- Sales, C. M. D., Gonçalves, C., 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. <https://doi.org/10.5920/mhldrp.2007.42145>
- Sales, C. M. D., Neves, I. T. D., Alves, P. C. G., & Ashworth, M. (2018). Capturing and missing the patient's story through outcome measures: A thematic comparison of patient-generated items in PSYCHLOPS with CORE-OM and PHQ-9. *Health Expectations*, 21(3), 615–619. <https://doi.org/10.1111/hex.12652>
- Sales, C. M. D., & Wakker, P. P. (2009). The metric-frequency measure of similarity for ill-structured data sets, with an application to family therapy. *British Journal of Mathematical and Statistical Psychology*, 62(3), 663–682. <https://doi.org/10.1348/000711008X376070>
- Sales, C. M. D., Wakker, P. P., Alves, P. C. G., & Fáisca, L. (2015). MF calculator: A web-based application for analyzing similarity. *Journal of Statistical Software*, 65(CODESNIPPET, 2, 1–10. <https://doi.org/10.18637/jss.v065.c02>
- Salmela-Aro, K. (1992). Struggling with self: The personal projects of students seeking psychological counselling. *Scandinavian Journal of Psychology*, 33(4), 330–338. <https://doi.org/10.1111/j.1467-9450.1992.tb00922.x>
- Salmela-Aro, K., & Nurmi, J.-E. (1996). Depressive symptoms and personal project appraisals: A cross-lagged longitudinal study. *Personality and Individual Differences*, 21(3), 373–381. [https://doi.org/10.1016/0191-8869\(96\)00078-5](https://doi.org/10.1016/0191-8869(96)00078-5)
- Sellen, J. L., Gobbett, M., & Campbell, J. (2013). Enhancing treatment engagement in sexual offenders: A pilot study to explore the utility of the Personal Aspirations and Concerns Inventory for Offenders (PACI-O). *Criminal Behaviour and Mental Health*, 23(3), 203–216. <https://doi.org/10.1002/cbm.1868>
- Sheffer, G., Canetti, L., & Wiseman, H. (2001). Psychometric properties of Goal-Attainment Scaling in the assessment of Mann's Time-Limited Psychotherapy. *Journal of Clinical Psychology*, 57(7), 971–979. <https://doi.org/10.1002/jclp.1063>
- Smith, A. (1994). Introduction and overview. In T. J. Kiresuk, A. Smith, & J. E. Cardillo (Eds.), *Goal attainment scaling: Applications, theory, and measurement* (pp. 1–14). Lawrence Erlbaum Associates.
- Tollefsen, T. K., Darrow, S. M., Lohne, V., & Berg-Nielsen, T. S. (2020). Experiences with using an idiographic assessment procedure in primary mental health care services for adolescents. *International Journal of Qualitative Studies on Health and Well-Being*, 15(1), 1763741. <https://doi.org/10.1080/17482631.2020.1763741>
- Tollefsen, T. K., Neumer, S.-P., & Berg-Nielsen, T. S. (2020). "What matters to you?": A randomized controlled effectiveness trial, using systematic idiographic assessment as an intervention to increase adolescents' perceived control of their mental health. *Journal of Adolescence*, 78, 53–61. <https://doi.org/10.1016/j.adolescence.2019.11.006>
- Urach, S., Gaasterland, C. M. W., Posch, M., Jilma, B., Roes, K., Rosenkranz, G., Van der Lee, J. H., & Ristl, R. (2018). Statistical analysis of Goal Attainment Scaling endpoints in randomised trials. *Statistical Methods in Medical Research*, 28(6), 1893–1910. <https://doi.org/10.1177/0962280218777896>
- Weisz, J. R., Chorpita, B. F., Frye, A., Ng, M. Y., Lau, N., Bearman, S. K., Ugueto, A. M., Langer, D. A., & 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. <https://doi.org/10.1037/a0023307>
- Wolpert, M., Cheng, H., & Deighton, J. (2015). Measurement Issues: Review of four patient reported outcome measures: SDQ, RCADS, C/ORS and GBO - their strengths and limitations for clinical use and service evaluation. *Child and Adolescent Mental Health*, 20(1), 63–70. <https://doi.org/10.1111/camh.12065>
- Wright, A. G. C., Beltz, A. M., Gates, K. M., Molenaar, P. C. M., & Simms, L. J. (2015). Examining the dynamic structure of daily internalizing and externalizing behavior at multiple levels of analysis. *Frontiers in Psychology*, 6, 1914. <https://doi.org/10.3389/fpsyg.2015.01914>

How to cite this article: Sales, C. M. D., Ashworth, M., Ayis, S., Barkham, M., Edbrooke-Childs, J., Fáisca, L., Jacob, J., Xu, D., & Cooper, M. (2022). Idiographic Patient Reported Outcome Measures (I-PROMs) for Routine Outcome Monitoring in Psychological Therapies: Position Paper. *Journal of Clinical Psychology*, 1–26. <https://doi.org/10.1002/jclp.23319>