Personal recovery self-report outcome measures in serious mental illness: A systematic review of measurement properties

Simon Felix, Kevin-Marc Valery, Meryl Caiada, Sarah Guionnet, Julien Bonilla-Guerrero, Jean-Marc Destaillats, Antoinette Prouteau

Clinical Psychology Review 112 (2024) 102459

Contents lists available at ScienceDirect

Clinical Psychology Review

journal homepage: www.elsevier.com/locate/clinpsychrev

Review

Personal recovery self-report outcome measures in serious mental illness: A systematic review of measurement properties

Simon Felix [a,b,*], Kevin-Marc Valery [a], Meryl Caiada [a], Sarah Guionnet [a], Julien Bonilla-Guerrero [b], Jean-Marc Destaillats [b], Antoinette Prouteau [a,b]

[a] Psychology Laboratory (LabPsy) UR4139, Bordeaux University, 3 ter place de la Victoire, 33000 Bordeaux, France
[b] ESPPAIR Unit, Department of Adult Psychiatry, Jonzac Hospital, 17500 Saint-Martial de Vitaterne, France

ARTICLE INFO

Keywords:
Personal recovery
Serious mental illness
Patient-reported outcome measures
Measurement properties
Psychometrics
Systematic review

ABSTRACT

Background: Personal recovery represents a paradigm shift in mental healthcare. Validated self-report outcome measures (PROMs) are needed to facilitate the transformation towards recovery-oriented practices and services. Objectives were to identify published measures and analyze their measurement properties using a standardized methodology.

Methods: Following the COSMIN guidelines, we conducted a systematic review of personal recovery PROMs in serious mental illness. The MEDLINE, PMC, PsycINFO, PsycARTICLES, PBSC and Scopus electronic databases were searched for articles published between May 2012 and February 2024. Full-text articles from a previous systematic review were also examined.

Results: 91 studies were included in the review, describing 25 PROMs. Ten of them had not been identified in previous reviews. Quality of evidence was globally poor for most PROM measurement properties. Very little evidence was found for cross-cultural validity, measurement invariance, measurement error and criterion validity. The Recovery Assessment Scale and Questionnaire about the Process of Recovery showed the strongest evidence for sufficient psychometric data on a wide range of measurement properties.

Conclusions: Several personal recovery measures are now available. While research is still needed to enhance their validity on some psychometric properties, the current tools appear sufficient to cover most research and clinical needs.

Recovery from serious mental illness (SMI) is a complex, highly debated issue that lacks consensus (Davidson et al., 2005; Leendertse et al., 2021). The medical model of mental illness defines recovery as an improvement of symptoms and other deficits associated with the disorder, and a return to a pre-existing healthy state (Davidson et al., 2008). However, persons with a lived experience of recovery from SMI have argued that this objective/category-based vision of recovery is not appropriate to mental illness (Andresen et al., 2011; Bassman, 2000; Crowley, 1997; Fisher, 1994; Tenney, 2000). They describe recovery as an individual journey rather than a common/uniform end state, that is not a question of returning to a premorbid state but rather incorporating the illness into a new sense of self (Hunter & Marsh, 1994). The mental health consumer movement thus supports a more subjective approach to recovery, known as a consumer-based, or personal definition of recovery: "a deeply personal, unique process […] a way of living a satisfying, hopeful, and contributing life even with limitations caused by illness" (Anthony, 1993, p.15).

Personal recovery has received growing interest over the past 20 years. Several attempts have been conducted to model personal recovery from SMI based on consumers' accounts. They have shed light on both the processes that are key to the recovery journey and the key steps common to most recovery paths (Andresen et al., 2011; Andresen et al., 2006; Andresen et al., 2003; Song & Shih, 2009). The CHIME framework (Leamy et al., 2011) synthesized the scientific literature pre-2011, thus offering a more consensual theoretical basis to conceptualize personal recovery, at least in its western-cultured meaning (Slade et al., 2014). Based on narrative synthesis analysis, the CHIME model proposes five non-linear steps of personal recovery, which simultaneously involve five processes: (1) Connectedness, (2) Hope and Optimism for the future, (3) Identity, (4) Meaning in life and (5) Empowerment.

Personal recovery thus represents a major paradigm shift for SMI care. The transformation of mental health services towards both person

* Corresponding author at: Laboratoire de Psychologie UR4139, 3ter place de la Victoire, 33000 Bordeaux, France.
E-mail address: simon.felix@u-bordeaux.fr (S. Felix).

https://doi.org/10.1016/j.cpr.2024.102459
Received 25 July 2023; Received in revised form 26 April 2024; Accepted 10 June 2024
Available online 13 June 2024

and recovery-centered care has therefore been defined as a priority by the World Health Organization (World Health Organization, 2021). Outcome measures of personal recovery could prove valuable to address the challenges that surround the notion of recovery. Personal recovery as a scientific concept itself would highly benefit from outcome measurement. Indeed, personal recovery models and frameworks rely mostly on qualitative data and expert opinions (Slade, Adams, & O'Hagan, 2012). There is a need to produce scientific evidence via standardized measures, in domains such as the longitudinal evaluation of the recovery process, or the link between advancing in the recovery journey and clinical domains (Slade et al., 2012). Measuring personal recovery is also critical to understand the cultural differences in play, and could allow its application to other populations, to determine specificities or communalities in the personal recovery process (Slade et al., 2014).

Moreover, Slade et al. (2014) argue that traditional clinical outcomes are not suited to assess these specific recovery processes. Therefore, to create a mental health system that prioritizes person-centered care, the authors insist that personal recovery outcome measures should be implemented in mental health care. Indeed, there is little data on the impact of mental health services on the recovery process (Slade, Leamy, et al., 2012). Additionally, the use of accurate outcome data in routine clinical practice is critical to the delivery of decent quality care (Newman-Taylor et al., 2019). Patient-reported outcome measures (PROMs) have been shown to inform different levels of the mental health system in terms of both efficiency, efficacy, and effectiveness of the recovery-oriented approach (Gordon et al., 2012). They not only reflect the subjective nature of recovery, but also promote the inclusion of user's perspectives in the process of health service provision (Krägeloh et al., 2015). And yet, there are important difficulties in creating recovery-oriented health services that consider users subjectivity and perspectives (Slade et al., 2014). Moreover, clinicians struggle to routinely use these outcome measures (Jensen-Doss et al., 2018; Newman-Taylor et al., 2019). A study investigating the routine use of standardized outcome measures amongst psychiatrists (Gilbody et al., 2002) showed that their non-use was justified by a lack of resources (not enough time or personnel), and critiques regarding their psychometric properties. In this regard, research is needed to provide appropriate validated assessment of personal recovery-related outcomes (Slade, 2010).

Personal recovery measurement has been exponentially investigated in the international literature. A key reference in the field was the document entitled 'Measuring the Promise: A Compendium of Recovery Measures Volume II' (Campbell-Orde et al., 2005). The authors had compiled in extensive detail all known measures of personal recovery, and described characteristics and psychometrics for each. Since then, there have been several literature reviews aimed at identifying tools to measure personal recovery (Burgess et al., 2011; Cavelti et al., 2012; Chirio-Espitalier et al., 2022; Law et al., 2012; Penas et al., 2019; Scheyett et al., 2013; Shanks et al., 2013; Sklar et al., 2013). A largely shared conclusion of these reviews, corroborated by the Leendertse et al. (2021) meta-analysis, is the current lack of a gold standard to measure personal recovery. However, previous reviews showed significant differences regarding their objectives and methodologies. Some were specifically focused on the assessment of personal recovery in psychotic disorders (Law et al., 2012), schizophrenia (Cavelti et al., 2012), or bipolar disorders (Chirio-Espitalier et al., 2022). Others aimed at identifying measures for routine use in health services, and therefore applied additional criteria of briefness and acceptability to consumers (Burgess et al., 2011; Penas et al., 2019). They also differed in inclusion criteria regarding measurement characteristics, such as self-report vs. clinician-rated measures or the assessment of recovery as an outcome vs. assessment of attitudes towards recovery, or styles of recovery. As a shared objective of these reviews was the exhaustive identification of recovery assessment instruments, a significant proportion did not limit their searches to peer-reviewed publication sources, and therefore described measures that had not been psychometrically tested. Additionally, a main limitation of previous reviews is the lack of a standardized procedure to analyze the psychometric properties of identified measures: when available, psychometric properties were described (Scheyett et al., 2013) or assessed with unstandardized criteria (Law et al., 2012; Penas et al., 2019; Sklar et al., 2013). Moreover, to our knowledge, no review has conducted any risk of bias and quality of evidence analyses, i.e., standardized assessment of methodological soundness, while it is essential for a reliable interpretation of psychometric properties.

The objectives of this systematic review were: (a) to identify personal recovery PROMs that have been the subject of a peer-reviewed publication, (b) to analyze their psychometric properties using a standardized methodology, and (c) to provide an overview of the psychometric quality of identified PROMs and formulate recommendations for their use and future research.

1. Methods

This systematic review followed the *Consensus-based Standards for the selection of health Measurement Instruments* (COSMIN) guidelines for the systematic review of Patient-Reported Outcome Measures (PROMs) (Mokkink et al., 2017; Prinsen et al., 2018). Appendix A. displays a list of abbreviations used throughout this article.

1.1. Search Strategy

1.1.1. Search

The search was conducted using a combination of free-text and controlled vocabulary search terms (MeSH terms, APA Thesaurus of Psychological Index Terms). It comprised four categories of terms: recovery-related (e.g., *recovery, personal recovery, recovery process,* etc.), assessment-related (e.g., *measure, scale, assessment,* etc.), mental health/illness-related (e.g., *mental health*, *serious mental illness, schizophrenia,* etc.), and COSMIN search filter terms. This last category refers to a search filter designed by the COSMIN authors to specifically identify studies on measurement properties. The filter has shown good sensitivity (Terwee et al., 2009). The four categories of terms were combined using the Boolean operator 'AND", and the terms inside each category were combined using the operator 'OR". When terms could be derived, we used the '*' operator to ensure better exhaustivity. This makes it possible to find all possible versions of a word (e.g., *assess** allows identification of studies containing *assessment, assessed, assessing* etc.).

A copy of the full term's list is available as supplementary material (document DS1. "*Search Strategy*"). The search focused on research conducted between May 2012 and February 2024. Related references of articles included in the review were also examined.

The search terms used for this review were extracted from the search strategy obtained from Shanks et al. (2013). This previous systematic review was used as a basis for this work, given its clear definition of personal recovery, limitation to SMI samples, and exhaustivity of search databases. The authors kindly provided us with their full list of reviewed studies, as well as their complete search strategy. As Shanks et al. ended their search in May 2012, we chose to limit our search strategy to articles published after this date. The original pool of $n = 35$ included studies from the Shanks review was thus screened and included if the studies met the eligibility criteria.

1.1.2. Electronic Databases

The following search engines were used to search in several databases: PubMed® was used to search the Medical Literature Analysis and Retrieval System Online (MEDLINE) and PubMed Central (PMC) databases; EBSCO*host* was used to search the Psychology and Behavioral Sciences Collection (PBSC), PsycINFO® and PsycARTICLES® databases (both were searched for optimal access to full-text articles); and the Scopus® integrated search engine for the Scopus® database.

1.2. Eligibility Criteria

Studies were included if they met the following criteria:

- Full text original research article (not including letter to editor, reviews, abstracts);
- Published in peer-reviewed journals;
- Published in English or French language;
- Published after May 2012;
- Utilized a patient-reported outcome measure (PROM) of personal recovery;
- Reported on at least one of the measure's psychometric properties (excluding aspects related to feasibility or interpretability);
- At least 75% of the sample was composed of people with SMI as defined by the National Institute of Mental Health (n.d.), aged between 18 and 65.

Studies were excluded if they met the following criteria:

- Sample composed of sex offenders, people with substance use disorders, or people with eating disorders
- Studies not aimed at PROM validation (named "PROM use" studies)

As recommended by the COSMIN guidelines, it was decided to exclude "PROM use" studies. These encompass studies that do not aim at PROM validation, only using PROMs as an outcome measurement instrument (e.g., PROMs used in randomized controlled trials, or in validation studies of other measurement instruments).

Notably, we only included published and peer-reviewed studies. Indeed, our goal was not to identify every existing personal recovery PROM, but rather to collect reliable psychometric data for the existing tools. Therefore, limiting to published peer-reviewed articles ensured a higher degree of quality and reliability of collected data.

1.3. Selection Process

The selection process for this study used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021).

Studies were extracted from the databases into a specialized systematic review software (Rayyan®). The provided automated tool was used to remove duplicates, set at 95% of similarities between studies. The remaining duplicates were removed by the first author (SF) to limit automation mistakes. Studies were first selected by title and abstract. The selection was performed independently by two authors (SF, MC). The computed agreement rate was 94.25%. In the event of disagreement, the study inclusion was discussed between the authors. Selected studies were sought for retrieval and obtained full texts were assessed for eligibility. If the eligibility criteria were met, studies were included in the review.

1.4. Data Extraction and Assessment

Data extraction and assessment were performed according to the COSMIN user manual recommendations and steps (Prinsen et al., 2018, Fig. 1).

1.4.1. Results of Studies on Measurement Properties

Psychometric data were extracted from each included study according to the COSMIN 'taxonomy of measurement properties' (Mokkink et al., 2010, Fig. 2). The following properties were extracted: structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypotheses testing for construct validity in terms of both convergent and known groups validity, and responsiveness.

The COSMIN 'updated criteria for good measurement properties' (Mokkink et al., 2018, pp. 28–29) were then used to rate the quality of reported measurement properties for each study as either *sufficient* (+), *insufficient* (−) or *indeterminate* (?).

As specified by the COSMIN guidelines, construct validity assessment required the review team to formulate hypotheses regarding the magnitude and direction of the expected effects. In this systematic review, the hypotheses were generated on the basis of the Leendertse et al. (2021) meta-analysis of personal recovery associated factors. We expected correlations >0.30 between PROMs and personal recovery related constructs described in the meta-analysis.

Responsiveness also required the review team to formulate a priori hypotheses, so for the purpose of this review, they were generated on the basis of the Thomas et al. (2018) meta-analysis of personal recovery longitudinal studies. We therefore expected that personal recovery PROMs should be able to detect small to medium-sized changes.

Criterion validity was also analyzed, although several authors (e.g., Leendertse et al., 2021) have underlined the lack of a gold standard to assess personal recovery. Indeed, COSMIN guidelines indicate that when a shortened version of a PROM is available, the longer version can be considered an adequate reference point for criterion validity analysis.

1.4.2. Risk of Bias Analysis

As recommended by the COSMIN guidelines, the methodological quality of the included studies was assessed using the COSMIN risk of bias checklist (Mokkink et al., 2017). For each measurement property assessed by a study, the checklist allows researchers to rate the methodological quality on a 4-point scale: *inadequate*, *doubtful*, *adequate,* or *very good*. Risk of bias is a tool rather than an objective in itself, as it is later used to grade the quality of evidence of a PROM's given measurement property, i.e., how much trust can be put on the conclusions.

1.5. Summary of Findings: Overall Rating and Quality of Evidence

For each identified PROM version, available psychometric data were qualitatively pooled by providing the range values for various parameters (range of model fit parameters on consistently found factor structure; range of Cronbach's α / McDonald's ω for internal consistency, etc.).

Qualitatively summarized results were rated using the COSMIN criteria for good measurement properties: the overall rating obtained could either be *sufficient* (+), *insufficient* (−), *indeterminate* (?), or *inconsistent* (±) when fewer than 75% of the studies for a PROM's property met the criteria.

The summarized results per measurement property per PROM were then graded regarding the overall quality of evidence (QoE). As recommended by COSMIN guidelines, the quality of evidence was graded as *high*, *moderate*, *low,* or *very low* based on an adapted version of the grading of recommendations assessment, development, and evaluation (GRADE) approach. It reflects the degree of confidence that the review team has regarding the estimation of the true measurement property. Of note, the review team decided that studies with inadequate risk of bias ratings would be excluded from the summarized results rating and the QoE analysis.

1.6. Supplementary Analysis: Content Validity of Included PROMs

Content validity is considered the most important psychometric property in the COSMIN guidelines, and is defined as "the degree to which the content of an instrument is an adequate reflection of the construct to be measured" (Mokkink et al., 2010). To provide information regarding content validity of included PROMs, we replicated the Shanks et al. (2013) method of mapping PROMs items onto the CHIME framework (Leamy et al., 2011). For PROMs that were freely available and/or for which items could be accessed, it was decided whether each item matched with one of the five CHIME dimensions: Connectedness, Hope, Identity, Meaning in Life and Empowerment. To properly define

what these dimensions theoretically encompass, we used the CHIME author's taxonomy of recovery processes (Leamy et al., 2011, Table DS2). Items were thus attributed to corresponding dimensions when possible or were rated indeterminate. Percentages of matching items per dimension and non-matching items were computed for each PROM. This analysis was done independently by SF and KMV, and an agreement rate was computed for each PROM. The mean agreement rate was 74.47%. Disagreements were addressed by discussion between the authors until a consensus was reached. This analysis allowed us to detail two aspects of content validity. Relevance (i.e., whether the PROMs items are relevant to the construct of interest within a specific population and context of use) was assessed as the percentage of items matching CHIME dimensions. We used the criteria of at least 85% relevant items described in Terwee et al. (2018). Comprehensiveness (i.e., no key aspects of the construct of interest should be missing) was assessed as whether all CHIME dimensions were covered by the PROMs items.

2. Results

2.1. Search Results

Flowchart is available as Appendix B. A total of 3031 articles were originally identified from our databases. The search was updated until February 22, 2024. Searching in the bibliography of included studies allowed the identification of 8 additional studies. The 35 articles included in Shanks' study were also reviewed. After eligibility assessment, the final number of included studies was $n = 91$.

Notably, Campbell-Orde et al. (2005) was excluded from the review despite its importance in the field, because of its non-peer-reviewed nature. Indeed, a large portion of this document's data originated directly from the authors who created the recovery PROMs, and these data were themselves often not published.

2.2. Included PROMs

A total of 25 personal recovery PROMs were identified from the 91 included studies. Including shortened or revised versions of these 25 scales, the total number of personal recovery PROMs rose to 34 measurement tools. A description of each included PROM is available in Appendix C.

Overall, we were able to identify three different types of personal recovery PROMs. The most common were outcome measures specifically designed to assess personal recovery (either as a process or as stages). We also identified PROMs that were originally designed to assess outcomes of recovery-promoting interventions: the Illness Management and Recovery Scale (Gingerich & Mueser, 2005), the West-Bridge Dual Recovery Inventory (Noel, Woods, Routhier, & Drake, 2016) and the Choice of Outcome for CBT in Psychosis Scale (Greenwood et al., 2010). Finally, we identified PROMs designed as indexes, i.e., composite scores computed from scales not originally designed to assess personal recovery. This was the case of the SAMHSA Recovery Inventory for Chinese (Chiu et al., 2010) and the Recovery Index (IsHak et al., 2017).

Of note, we did not include the Indonesian Recovery Scale for Patients with Schizophrenia (Saputra et al., 2022) as its recent publication did not provide any data on psychometric properties of interest. The Eating-Disorders Recovery Questionnaire (Bachner-Melman et al., 2021) and Vietnamese Mental Health Recovery Scale Lim, Byrne, Shieh, Hồ, and Mason (2017) were excluded as they did not meet the sample characteristics criteria. The Short Interview to assess STages of Recovery (Wolstencroft et al., 2010) was also excluded, as it is not self-rated and therefore was not considered as a PROM.

2.3. Characteristics of Included Studies

A description of the characteristics of the included studies is available as supplementary material (Table S2. *Characteristics of the Included Studies*). The complete reference list of included studies is also available as supplementary (Document DS2. *Reference List of Included Studies*).

The studies were very heterogeneous in terms of number of participants (n ranging from 35 to 4041), nature of SMI (schizophrenia spectrum disorders, bipolar disorders, depressive disorders, personality disorders, anxiety disorders, etc.) and mean ages of participants (21 to 55 years old). Most studies were conducted in the United States ($n = 16$), United Kingdom ($n = 17$), Australia ($n = 9$), the Netherlands ($n = 8$), and Hong-Kong/China ($n = 10$). English was by far the most used PROM language ($n = 45$).

2.4. Results of Studies on Measurement Properties

The table reporting raw psychometric data for each included study with the associated risk of bias ratings is available as supplementary material (Table S1. *Results of Studies on Measurement Properties*).

2.5. Summary of Findings

Summarized results and quality of evidence ratings are available as supplementary material (document *DS3. Summary of Findings*). The document presents a summarized description and analysis of collected psychometric properties in table and text format. The subscales and domains supposedly assessed by the PROMs are displayed in the tables.

2.6. Summarized Results

Table 1 summarizes the available psychometric data of the included PROMs and their associated Quality of Evidence ratings.

2.6.1. Content Validity Analysis

The mapping of PROMs items onto the CHIME framework is available in Appendix D. Regarding the relevance of included PROMs items, nine measures (out of 25) were found to yield <85% of items mapping to the CHIME dimensions: the MHRM-30 (83.3%), MHRM-10 (70%), STORI-50 (60%), POP-RS-13 (53.8%), PRI-25 (8%), BRQ-36 (66.7%), RMQ-24 (83.3%), SISR-A&B (33.3%) and SeRvE-40 (82.5%). The Psychosis Recovery Inventory (PRI, Chen et al., 2005) was found to have poor items' relevance regarding the CHIME dimensions, with almost all items considered not to map the framework ($n = 23$, 92% non-mapping items). Indeed, items were very oriented towards symptom recovery from a first psychotic episode. This typifies what Andresen et al. (2011) described as one of the many meanings that can be given to the term recovery in the literature: a meaning based on the medical model, referring to the end of the psychotic phase of a discrete episode of schizophrenia. However, this PROM had been included in several previous reviews of personal recovery measures (Cavelti et al., 2012; Law et al., 2012; Scheyett et al., 2013; Shanks et al., 2013). Given this element of content validity, we would not recommend using the PRI as a personal recovery PROM.

Regarding the comprehensiveness of included PROMs, seven measures were found to not encompass every CHIME dimension: MHRM-10 (no items for Connectedness, Hope, Empowerment), POP-RS-13 (no items for Connectedness, Identity, Meaning in Life), PRI (no items for Connectedness, Hope, Identity and Meaning in Life), BRQ-36 (no items for Hope), SISRA&B (no items for Connectedness and Hope), HAO (no items for Identity) and CRM-15 (no items for Identity).

Table 1
Summarized Results of PROMs Measurement Properties and Associated Quality of Evidence Ratings.

PROM version	Structural validity	Internal Consistency	Cross-cultural validity / Measurement invariance	Reliability of total score	Reliability of subscales	Measurement error	Criterion validity	Hypotheses testing for Convergent validity	Hypotheses testing for Known-groups / Discriminative validity	Responsiveness
RAS										
41 items	(-) Low	–	–	(+) Very low	–	–	–	(+) Moderate	(+) Low	–
24 items	(+) Low	(+) Very low	(+) Low	(+) Moderate	(+) Moderate	–	–	(+) Moderate	–	(+) Very low
QPR										
22 items	(+) Moderate	(-) Low	–	(+) Low	(+) Moderate	–	–	(+) Moderate	(+) Low	(+) Low
15 items	(+) High	(+) High	–	(+) Moderate	–	(?) Low	–	(+) High	–	(+) Moderate
MHRM										
30 items	(?) Moderate	–	–	–	–	(-) Very low	–	(+) High	–	(-) Very low
10 items	–	–	–	(+) Very low	–	–	–	(+) High	(+) Moderate	–
STORI										
50 items	–	–	–	–	–	–	–	(+) Moderate	–	–
30 items	(?) Low	–	–	–	(+) Very low	–	–	(+) Moderate	–	–
IMRS-Client										
15 items	(+) Very low	(-) Very low	–	(+) Moderate	(-) Very low	(-) Very low	–	(±)	–	(-) Very low
POP-RS										
13 items	(?) High	–	–	–	–	–	–	(+) High	(-) Low	–
RAS-DS										
38 items	(+) High	(+) High	–	–	–	–	–	(+) High	–	–
20 items	(+) High	(+) High	–	–	–	–	–	(+) High	–	–
I.ROC										
12 items	(+) Low	(+) Very low	(+) Low	(+) High	–	–	–	(+) High	–	(-) Low
SubRAS										
17 items	–	–	–	(+) Very low	–	–	–	(+) High	–	–
PRI										
25 items	–	–	–	(-) Very low	(+) Very low	–	–	(-) Moderate	(-) Low	–
RPI										
22 items	(?) Moderate	–	–	–	(-) Low	–	–	(+) Very low	–	–
MARS										
25 items	(+) High	(+) High	–	(+) Low	–	–	–	(+) Moderate	–	–
MHRS										
10 items	(?) Moderate	–	–	–	–	–	–	(-) Low	–	–
RI										
21 items	(?) Moderate	–	–	–	–	–	–	–	–	–
WBDRI										
14 items	–	–	–	–	–	–	–	–	–	–
SAMHSA-RIC										
41 items	(-) Low	–	–	–	–	–	–	–	–	–
72 items	–	–	–	–	–	–	–	–	–	–
BRQ										
36 items	–	–	–	(+) Very low	–	–	–	(+) Very low	–	–
MVML										
17 items	–	–	–	–	–	–	–	–	–	–
RMQ										
24 items	(+) High	(+) High	–	–	–	–	–	–	–	–
CHOICE										
24 items	(?) Low	–	–	–	(+) Very low	–	–	(+) Low	–	–
11 items	(+) High	(+) Moderate	–	–	–	–	–	(+) Very low	–	–
SISR-A										
1 item	–	–	–	(-) Very low	–	–	–	(+) Low	–	–
SISR-B										
4 items	–	–	–	(-) Very low	–	–	–	(+) High	–	–
SRS										
45 items	(?) Moderate	–	–	(+) Very low	–	–	–	(+) Moderate	(±)	–
SeRvE										
40 items	–	–	–	–	–	–	–	(+) Low	–	–
15 items	–	–	–	–	–	–	–	(+) Low	–	–
HAO										
4 items	(+) Very low	(-) Very low	–	(-) Low	–	–	–	(+) High	–	–
CRM										
15 items	(+) High	(+) High	–	–	–	–	–	–	–	–

Note. Properties quality ratings: (+) = Sufficient; (−) = Insufficient; (?) = Indeterminate; (±) = Inconsistent. Color code: pale green = Sufficient property with Low QoE, fern green = Sufficient property with Moderate QoE, bright green = Sufficient property with High QoE. Quality of evidence ratings range from Very low to High. Dashes indicate missing data. PROM = Patient-Reported Outcome Measure; RAS = Recovery Assessment Scale (RAS); QPR = Questionnaire about the Process of Recovery; MHRM = Mental Health Recovery Measure; STORI = STages Of Recovery Instrument; IMRS-C = Illness Management

and Recovery Scale -Client version; POP-RS = Peer Outcome Protocol – Recovery Scale; RAS-DS = Recovery Assessment Scale – Domains and Stages; I.ROC = Individual Recovery Outcomes Counter; SubRAS = Subjective Recovery Assessment Scale; PRI = Psychosis Recovery Inventory; RPI = Recovery Process Inventory; MARS = Maryland Assessment of Recovery Scale; MHRS = Mental Health Recovery Star; RI = Recovery Index; WBDRI = WestBridge Dual Recovery Inventory; SAMHSA-RIC = SAMHSA-Recovery Inventory for Chinese; BRQ = Bipolar Recovery Questionnaire; MVML = My Voice My Life; RMQ = Recovery Markers Questionnaire; CHOICE = CHOIce of outcome for CBT in psychoses; SISR-A = Self-Identified Stage of Recovery – part A; SISR-B = Self-Identified Stage of Recovery – part B; SRS = Stages of Recovery Scale; SeRvE = Service-user Recovery Evaluation scale; HAO = Hope Agency and Opportunity; CRM = Consumer Recovery Measure.

3. Discussion

3.1. Main Results

This review aimed to identify personal recovery PROMs and analyze their psychometric properties. Its main strength was to propose a thorough systematic investigation of published PROMs, using standardized criteria to evaluate both (a) psychometric properties of included measures and (b) quality of evidence. In the peer-reviewed literature, we identified 25 PROMs designed to assess personal recovery in SMI. Two PROMs stood out significantly in terms of both (a) number of included studies and (b) overall psychometric quality: the Recovery Assessment Scale (RAS) and the Questionnaire about the Process of Recovery (QPR). Three other PROMs showed promising quality of psychometric data, though requiring further investigation: the Recovery Assessment Scale - Domains and Stages (RAS-DS), the Individual Recovery Outcomes Counter (I.ROC) and the Maryland Assessment of Recovery Scale (MARS). These scales are discussed below.

Notably, 10 of the 25 identified PROMs had never been included in any of the previously identified reviews: the Recovery Assessment Scale – Domains and Stages (RAS-DS, Hancock et al., 2015), Subjective Recovery Assessment Scale (SubRAS, Yildiz et al., 2017), Recovery Index (RI, IsHak et al., 2017), WestBridge Dual Recovery Inventory (WBDRI, Noel et al., 2016), SAMHSA-Recovery Inventory for Chinese (SAMHSA-RIC, Chiu et al., 2010), MyVoiceMyLife (MVML, Gordon et al., 2012), CHoice of Outcomes for CBT in psychosEs (CHOICE, Greenwood et al., 2010), Service User Recovery Evaluation scale (SeRvE, Barber et al., 2012), Consumer Recovery Measure (CRM, Lusczakoski et al., 2016) and the Hope, Agency and Opportunity (HAO, Newman-Taylor et al., 2017).

3.1.1. RAS-24

The Recovery Assessment Scale-24 items (RAS-24, Corrigan et al., 2004) is a short version of the original RAS-41. According to Campbell-Orde et al. (2005), the RAS-41 was originally designed as an outcome measure for program evaluation (Giffort et al., 1995). The items were generated upon analysis of four consumers' recovery stories. We do not recommend the use of the 41-item version, as it severely lacks psychometric testing. In order to develop the short RAS-24 version, Corrigan et al. (2004) deleted three of the eight components (originally identified using Principal Component Analysis, PCA), which led to a good fitting solution with confirmatory factor analysis (CFA). The remaining 24 items cover five dimensions: *Personal confidence and hope, Willingness to ask for help, Goal and success orientation*, *Reliance on others* and *Lack of domination by symptoms*. The RAS-24 was by far the most used PROM in our study pool ($n = 18$ studies). It showed sufficient ratings on a wide variety of psychometric properties, though lacking quality of evidence: structural validity for its 5-domain structure, internal consistency, cross-cultural validity, reliability of both total and subscale scores, convergent validity and responsiveness. Regarding content validity, the RAS-24 had good relevance and comprehensiveness. The RAS-24 items were found to evenly encompass the five CHIME dimensions, whereas the 41-item version had a strong focus on illness self-management. Two items did not map any of the five dimensions: item 16 ('*My symptoms interfere less and less with my life*") and item 17 ('*My symptoms seem to be a problem for shorter periods of time each time they occur*"). The main limitation of the RAS-24 was the globally limited quality of evidence of its psychometric properties. This was mainly due to inconsistency across studies. No data was available regarding the RAS-24 known-groups validity, measurement error and criterion validity. Although we found evidence for sufficient validity of the RAS-24 five-factor structure, we recommend that a new exploratory factor analysis (EFA) be conducted, as the original five-factor structure originated from the debatable item reduction procedure described above. This might explain the inconsistency we found for both the structural validity and internal consistency of this structure. Additionally, the item reduction might have impacted the content validity of the scale, as it was neither motivated by theoretical elements nor decided with consumer feedback (at least to our knowledge). Moreover, consumers have argued that the RAS-24 failed to capture components of the later stages of the recovery journey (Hancock et al., 2015; Hancock et al., 2013).

3.1.2. QPR-22

The Questionnaire about the Process of Recovery (QPR, Neil et al., 2009) is a 22-item Likert-type scale developed in collaboration with service users. Items were generated on the basis of the Pitt et al. (2007) study, a thematic synthesis of recovery accounts of people with psychotic disorders that identified three key themes of recovery: rebuilding self, rebuilding life and hope for a better future. The QPR-22 underwent quite thorough psychometric investigation, with $n = 6$ included studies. It supposedly assesses two dimensions of personal recovery, with an *Interpersonal* and an *Intrapersonal* factor. It showed sufficient ratings for structural validity with this two-factor solution, reliability for total and subscales scores, convergent and known-groups validity, and responsiveness. Regarding content validity, the QPR-22 showed good relevance and comprehensiveness, with a good fit to the CHIME framework and items equally covering all five dimensions. No unmatching items were found. No data was identified regarding QPR-22 cross-cultural validity/measurement invariance, measurement error and criterion validity. It received an insufficient rating for internal consistency as its *Interpersonal* factor failed to reach satisfactory values in three out of four studies. Although it received a sufficient rating for structural validity, the QPR-22 two-factor structure was found to be inconsistent across studies and lacked coherence with its theoretical basis. Notably, a 3-factor solution more coherent to the scale's theoretical basis had been identified by Chien and Chan (2013) using PCA, but was never replicated. The 15-item version (see below) addresses some of the scale's limitations.

3.1.3. QPR-15

The QPR-15 was developed as a shorter version of the QPR-22 to address structural validity and internal consistency limitations of the longer version. Law et al. (2014) removed the 7 items from the problematic *Interpersonal* factor. The scale is therefore considered unidimensional. The QPR-15 yielded several sufficient measurement properties with overall satisfying quality of evidence: structural validity for the PROM's unidimensionality, internal consistency, reliability for total score, convergent validity and responsiveness. Regarding content validity, the QPR-15 also had good relevance and comprehensiveness. As with its longer version, the QPR-15 items adequately fit the CHIME framework, with a small emphasis on Meaning in Life statements. No unmatching items were found. Regarding the scale's limitations, although one study investigated measurement error, it did not provide enough information and the result was rated indeterminate. No data regarding cross-cultural validity/measurement invariance, criterion

validity and known-groups validity was found. As with the RAS-24, the item reduction procedure that led to the QPR-15 is debatable, and might have impacted content validity of the scale as it was done without consumer feedback and seemingly solely for statistical purposes. We found only one study that tested the structural validity of the QPR-15 using EFA (Law et al., 2014), which yielded a single-factor solution with appropriate fit. We suggest that replication of the single-factor structure using EFA is needed to ensure the relevance of the reported CFA results.

3.1.4. RAS-DS-38

The Recovery Assessment Scale – Domains and Stages (RAS-DS, Hancock et al., 2015) is a 38-item PROM based on the RAS and is specifically designed to address its imperfections. With the help of consumers, ill-fitting and redundant items were removed, and new items were introduced that assess the later stages of recovery. The scale assesses four dimensions: *Doing things I value*, *Looking forward*, *Mastering my illness*, *Connecting and belonging*. RAS-DS-38 underwent few validation studies ($n = 4$) but obtained sufficient ratings for structural validity, internal consistency and convergent validity, all with high QoE. Regarding content validity, the scale showed good relevance and comprehensiveness. The 38 items of the RAS-DS covered all CHIME domains, although with a strong emphasis on the Empowerment dimension ($n = 14$ items, 36.8% of all items). This version appears to be particularly relevant for the assessment of the later stages of personal recovery. Indeed, it has previously been suggested that empowerment is important in the later stages of personal recovery, from fostering the determination to realize the goal of recovery, to the process of regaining control over illness and life (Andresen et al., 2011). Of note, two items were not considered to map the CHIME framework: items 5 ('*I do things that are valuable and helpful to others*") and 24 ('*I can learn from my mistakes*"). Regarding the scale's limitations, no data was found regarding the RAS-DS-38 cross-cultural validity / measurement invariance, reliability of both total and subscale scores, measurement error, criterion validity, known-groups validity and responsiveness. With data on reliability and responsiveness lacking, the scale cannot reliably and validly assess changes in personal recovery over time. Although it obtained high QoE ratings, this result should be nuanced as it might originate from limitations inherent to COSMIN methodology (see the later 'Limitations Regarding COSMIN Methodology" section for details). Further psychometric testing is required, especially regarding the structural validity of the scale. Of note, a recent shorter version of the scale (RAS-SF-20, De Silva et al., 2021) showed better structural validity but lacks replication (n = 1 included study).

3.1.5. I.ROC-12

The Individual Recovery Outcomes Counter (I.ROC, Monger et al., 2013) is a 12-item scale designed in Scotland by a mental health organization called Penumbra. It was created to measure service users' progress in 12 key areas of recovery identified by a working group of senior professionals. The measure was developed in collaboration with service users and was pilot-tested for comprehensibility. It supposedly assesses two dimensions of recovery: *Intrapersonal* and *Interpersonal*. The I.ROC-12 was used in $n = 5$ validation studies. It yielded sufficient reliability for the total score, as well as sufficient convergent validity, both with high QoE. It was one of the two only measures to receive a sufficient rating for measurement invariance. Evidence of sufficient structural validity was found for a two-factor structure, when response categories were collapsed from six to four response options. Internal consistency was also rated sufficient. We could not assess the I.ROC-12 content validity. Indeed, the I.ROC-12 is copyrighted. We were therefore not able to access it in time for this review. Again, no data was found regarding several psychometric domains: measurement error, criterion validity and known-groups validity. The I.ROC-12 suffers several limitations regarding psychometry, notably with poor QoE for structural validity and internal consistency owing to high inconsistency across studies. This inconsistency could be due to the number of response categories, as both Rudd et al. (2020) and Sportel et al. (2023) did not specify whether they used the six- or four-response option forms. We also found the I.ROC-12 responsiveness to be insufficient, with the available studies mainly rated inadequate, and one adequate study showing too small correlations with change scores of other outcome instruments (Sportel et al., 2023).

3.1.6. MARS-25

The Maryland Assessment of Recovery Scale (MARS, Drapalski et al., 2012) was developed by a team of clinical scientists in collaboration with experts and consumers. Its items were derived from a revised version of the Substance Abuse and Mental Health Services Administration (SAMHSA) operational definition of recovery that was developed by the team. The PROM was pilot-tested and checked for comprehensibility. The first validation study in people with SMI led to the 25-item version. The MARS-25 was involved in three validation studies. Sufficient structural validity was found for a unidimensional factor structure, as well as sufficient internal consistency, test-retest reliability and convergent validity. Regarding content validity, the MARS-25 fitted the CHIME framework well, with only one item not considered to match any dimension: item 20 ('*I know that I can make changes in my life even though I have a mental illness*"). However, we found that it placed a strong emphasis on the Empowerment dimension ($n = 11$ items, 44% of total items) at the expense of the Connectedness and Meaning in Life dimensions. Nevertheless, all CHIME dimensions were represented across the items. Regarding the scale's limitations, no data was found for cross-cultural validity/measurement invariance, measurement error, criterion validity, known-groups validity and responsiveness. Further research is thus required. The current lack of responsiveness data hinders the scale from being used for longitudinal assessment.

3.1.7. Other Widely Used Scales

Two other scales, the Mental Health Recovery Measure (MHRM) and the Illness and Management Recovery Scale - Client version (IMRS-C), have received much attention in the literature. However, the present review showed that their validation is presently limited.

Although the MHRM-30 had interesting content validity characteristics, the MHRM-10 showed poor comprehensiveness (no items for the Connectedness, Hope and Empowerment dimensions). Both the 30- and 10-item versions did not present any conclusive evidence for several important construct validity and reliability parameters. No clear factor structure could be identified, as previous EFAs did not report fit parameters and no CFA was conducted. We recommend that a clear and strong evidence base for the MHRM structural validity be developed before carrying out a CFA (Brown, 2015).

In our opinion, the IMRS-C-15 should be used with caution as a broad personal recovery outcome measure, as it was developed to assess specific outcomes related to the Illness Management and Recovery program (IMR, for a recent review of effectiveness see Goh et al., 2023). Notably, its items covered all of the CHIME dimensions. Additionally, our results suggest that the IMRS-C-15 suffers from issues regarding internal consistency (at least for the three-factor structure), reliability of subscales and measurement error. It received an insufficient rating for responsiveness.

3.2. Limitations

3.2.1. Current Validity Limitations of Personal Recovery PROMs

In the present review, three measurement properties were found to

be very rarely tested: criterion validity, measurement error and cross-cultural validity/measurement invariance.

Although we defined no gold-standard for personal recovery assessment, we found very few studies assessing criterion validity. Studies that compared PROM short version scores with the long version scores were all rated inadequate in the risk of bias analysis. Indeed, they had computed the short version scores from the participants' answers to the long version PROM, logically leading to high correlations.

Regarding measurement error, only two of the 91 included studies computed the minimal important change score, without providing the smallest detectable change value nor the limits of agreement. Further research is needed in this area, as measurement error is a key parameter to interpret a scale's change score and is therefore crucial for recovery-oriented practice assessment.

Finally, we found very few studies that assessed cross-cultural validity or measurement invariance. Leendertse et al. (2021) highlighted the scarcity of studies investigating the effects of socio-demographics and patient characteristics on personal recovery: they reported associations with type of diagnosis and physical health. These associations could originate from variance in the measures used. Moreover, Llewellyn-Beardsley et al. (2023) recently stressed the importance of structural barriers that could prove detrimental to the recovery process at the socio-economical, institutional, and political level, and which are often overshadowed by the individualistic emphasis of most recovery narratives. The psychometric functioning of recovery measures should be explored within these different contexts.

However, the lack of data on cross-cultural validity, measurement invariance and measurement error is not specific to the personal recovery construct. A quick overview of recently published COSMIN reviews yielded similar results (e.g., empathy measurement, Harrison et al., 2022; workplace inclusion, Rezai et al., 2020; health-related quality of life in patients with lymphedema, Beelen et al., 2021; etc.). This could be due to the cost of these psychometrics evaluations in terms of time and participant resources.

Despite the large number of identified PROMs, only a few were generated on the basis of published conceptualizations/models of personal recovery. The QPR was based on the Pitt et al. (2007) model of personal recovery, the MHRM was based on the Young and Ensing (1999) model, and the STORI and SISR were based on the Andresen et al. (2003) model. The MARS and SAMHSA-RIC were based on the SAMHSA operational definition of personal recovery. We found no measure that used the CHIME framework as a basis to generate items, although it is recognized as somewhat consensual amongst recovery experts (Song, 2017). Of note, the INSPIRE scale (Williams et al., 2015) was based on CHIME, but was designed as a measure of staff support for personal recovery. The scale was not included in the present review as it is not a personal recovery PROM.

3.2.2. Limitations Inherent in the Present Study

This study presents several limits. First, using Shanks et al. (2013) review as a basis for updating the literature search, and therefore only including studies published after May 2012, is a limitation to the systematic aspect of the review. However, if a study appeared in our search that was published before May 2012, we still considered it for inclusion. Additionally, we analyzed the references of included studies, in order to ensure better exhaustivity. This led to the inclusion of two studies published before May 2012 that had not been identified earlier. We also chose to include only peer-reviewed articles, to ensure a minimal quality of included studies, and to avoid methodological bias as much as possible. However, this may have prevented us from identifying non-validated PROMs. Furthermore, as noted by Shanks et al. (2013), the exclusion of studies using samples mainly composed of people with eating disorders and substance use disorders might have led to missing some other PROMs. Nonetheless, the ones we identified are cited in the Results section of this review. Subsequently, our decision to exclude "PROM use" studies (i.e., not aimed at PROM validation) may have led to disregard available data on psychometric properties such as responsiveness (e.g., from randomized controlled trials) or convergent validity (e.g., from transversal correlational studies). Yet, the COSMIN guidelines explicitly recommend to apply this exclusion criterion, for feasibility and exhaustivity reasons (p.20 of the COSMIN user manual).

Another limitation of this study is that we decided to only discuss the PROMs with the most promising psychometrics. Nevertheless, each PROM has its own benefits and limitations, and could be useful for a specific purpose in a specific context. For example, the SeRvE scale (Barber et al., 2012) is the only included PROM with a specific focus on spiritual well-being. Likewise, some PROMs were designed for contexts where time constraints are a priority (e.g., SISR-A&B, HAO-4).

Finally, the method of mapping PROMs items onto the CHIME framework as a content validity analysis is also debatable. This method was efficient in identifying biomedically oriented items (i.e., theoretically not relevant for personal recovery), as well as in verifying that the PROMs items covered all dimensions of the CHIME framework evenly. However, as seen above, the CHIME framework is not the only model of personal recovery available. Moreover, the CHIME dimensions and subdimensions express key themes of personal recovery rather than definitions of specific constructs or processes. This may have introduced a bias in the mapping process, such as inter-rater variability. To limit this bias, the present mapping was conducted independently by two authors and disagreements were discussed until consensus was reached.

3.2.3. Limitations Regarding COSMIN Methodology

As observed by others (e.g., Meinck et al., 2022), the COSMIN methodology may sometimes be too severe? The 'worst score counts' principle generally leads to a high risk of bias and low QoE ratings. As underlined by Carlton et al. (2022), there is also an effect of publication year: the quality standards for PROM validation articles were not the same 10 years ago as they are today, and this often resulted in poorer ratings for older studies.

Second, we found the COSMIN methodology to lack finesse on several points. Regarding the construct validity and responsiveness summarized data, we found the percentage of confirmed hypotheses to poorly reflect PROMs' performance. A weighting computation considering the total number of tested hypotheses may be a better alternative. Similarly, the QoE rating does not factor in the total number of included studies for a specific PROM. This led to an unwanted bias: the most studied PROMs often yielded some degree of inconsistency amongst studies, thus obtaining lower QoE ratings, whereas they should benefit from this larger scientific data basis. On the other hand, PROMs used in fewer studies received higher QoE ratings when their methodological quality was good. We took this limitation into account to formulate our recommendations of use. Nevertheless, we consider that future guidelines of the COSMIN methodology should integrate this limitation, as the imprecision criterion that allows the QoE rating to be penalized in the event of a small total sample size is not sufficient to account for the effect of the total number of studies.

We also identified limitations regarding the COSMIN criteria for good measurement properties (pp.28–29 of the COSMIN user manual). Regarding Classical Test Theory (CTT) indicators of structural validity supported by the COSMIN guidelines, some evidence suggests that they could lead to improper conclusions. Indeed, as explained in Brown (2015), Hu and Bentler (1999) criteria for establishing adequate

structural validity should not be used in a categorical manner, i.e., CFI > 0.95 is good and < 0.95 is insufficient ("as they fluctuate as a function of modeling conditions"). The different fit parameters should be compared with each other and should be used in tandem with other aspects of the fitted solution, such as the localized areas of strain. The chosen COSMIN criteria led to sufficient structural validity ratings for studies in which the confirmatory factor analysis solution simultaneously yielded a CFI < 0.70 (which is a largely insufficient value) and an RMSEA <0.06, a value that COSMIN considers sufficient. A combination of less severe fit criteria is preferable to the current mutually exclusive propositions. We also found specific methodological limitations regarding the assessment of measurement error: the user manual does not provide clear guidelines for calculating the limits of agreement or the smallest detectable change values. We based our calculi on the COSMIN recommended de Vet et al. (2011) book, but other methods could have been used. Regarding the rating of collected structural validity data, it was decided in this review to rate studies that used PCA for structural validity assessment as inadequate. Indeed, we considered that there is sufficient published evidence that this method is inappropriate (Gruijters, 2020, Fabrigar et al., 1999) for single outcome scales validation, and the COSMIN criteria for good measurement properties do not offer specific criteria to rate PCA results. However, this might be considered too strict.

Finally, if studies produce inconsistent results regarding the psychometric properties of a PROM, the guidelines could provide the review team with standardized criteria to help identify a precise source of inconsistency. In the current document version, this whole process is entirely up to the review team, so this issue would benefit from a modification.

3.3. Perspectives For Future Research

This review underlines the large number and heterogeneity of personal recovery PROMs in the current literature. Although the Recovery Assessment Scale (RAS) and Questionnaire about the Process of Recovery (QPR) stood out in overall quantity of tested psychometric properties, research is still needed to address some of the limitations described above. Both scales would benefit from an in-depth content validity assessment, to address the potential limitations due to their shortening. However, available psychometric evidence for these two PROMs is sufficient to make them currently the best candidates for criterion validity analysis, although they have not been sufficiently validated to be strictly considered as gold standards. The RAS lacks clear evidence for its structural validity and has been criticized by consumers for not being sufficiently representative of the later stages of recovery (Hancock et al., 2013, 2015), so until the RAS-DS-38 has received further validation, we recommend using the QPR for this purpose.

On a purely methodological note, the reader's attention should be drawn to the process of developing a shorter PROM version: aiming to develop shorter PROMs should never be done at the expense of content validity. In this review, numerous studies proceeded to item reduction as a way to improve structural validity. However, as pointed out by Koğar (2020), item reduction should be undertaken primarily for reasons of feasibility, not by the need to improve indicators of structural validity. For example, in CTT analysis, various techniques were identified in the included studies: deleting all the items from a factor that explains little of the total variance; arbitrarily keeping the items with the highest loadings; or deleting items upon inspection of localized fit indices. In IRT analysis, infit and outfit values were sometimes used to delete ill-fitting items prior to conducting a CFA. These techniques are not recommended as best practice for item reduction, especially when they are not theory-driven (for a review, see Boateng et al., 2018).

We also urge researchers to limit the creation of new personal recovery PROMs, but rather to use the available ones with stronger evidence, translating and adapting them to their specific context if needed (for guidelines, see International Test Commission, 2017; *American Educational Research Association, American Psychological Association*, 2014; or Gana et al., 2021 for French-speaking readers). This should ensure the subsequent psychometric enhancement of previously developed PROMs and make studies more comparable. Indeed, the heterogeneity between personal recovery measures has been found to impede the comparability of results (Leendertse et al., 2021).

Further research is also needed to probe the content validity characteristics of personal recovery PROMs, beyond our analysis of relevance and comprehensiveness of items. PROMs should meet several criteria when they are designed, such as involving people with mental illness, using a theoretical conceptualization of recovery, being pilot tested for clarity and comprehensibility, meeting reasonable standards of acceptability and feasibility. Otherwise, they may never be used for routine assessment or research purposes. This specific point could be the subject of another fully fledged systematic review, with the recently developed COSMIN checklist for PROM content validity analysis (Terwee et al., 2018).

4. Conclusion

This systematic review showed the RAS and QPR to be the most valid personal recovery PROMs, although the QPR-15 has better structural validity and internal consistency. However, none of the included PROMs had sufficient properties on all psychometric domains with high levels of evidence. Further research is needed to extend knowledge on the measurement characteristics of PROMs. Most of the studies included were conducted in the U.S.A., U.K., and Australia. Although the recovery movement has long started propagating internationally, this result underlines these countries' cultural, societal, and political advance in the provision of recovery-oriented health care. Thanks to knowledge gained in the field and dedicated funding, this position of leadership has allowed researchers to engage in developing validated personal recovery PROMs, which in turn helps in organizational matters at all levels (Slade et al., 2010). As shown in this study, the currently available database of PROMs is well endowed, although most of them still need further research to enhance their psychometric properties. Nevertheless, the currently available PROMs appear sufficient to cover the needs of personal recovery assessment, whether for routine clinical practice or research purposes. Personal recovery PROMs are useful as mental health care indicators and are a step in the right direction for the global transformation of mental health organizations as they seek to improve the pathways to personal recovery (Anthony, 2000; Roe et al., 2022; Slade, 2010).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Ray Cooke for copyediting this article, as well as Hélène Plouseau-Guédé for her precious help in developing the search strategy.

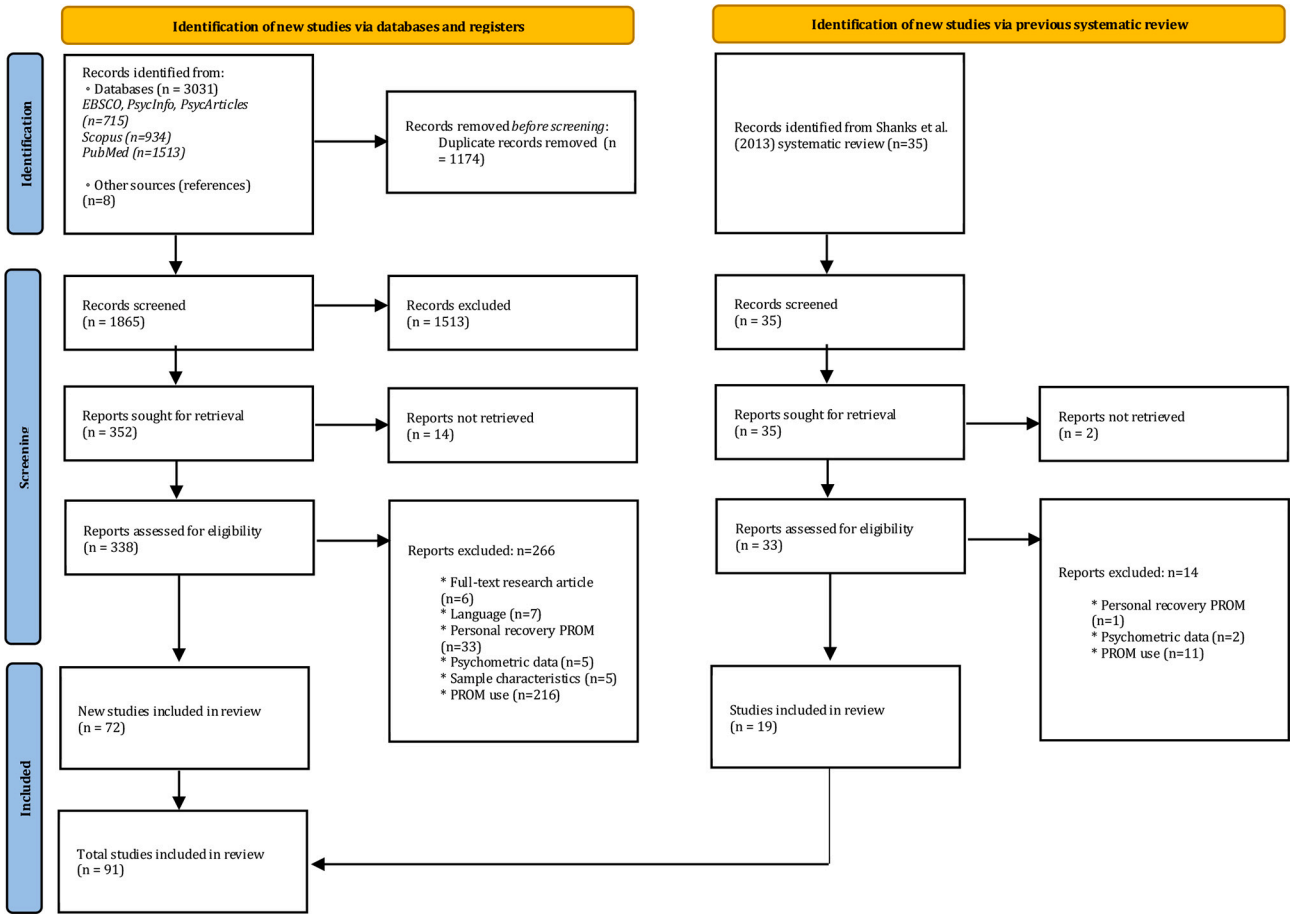
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Abbreviations

Common-term abbreviation	Full-term
CBT	Cognitive Behavioral Therapy
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CHIME	Connectedness, Hope, Identity, Meaning in Life and Empowerment
COSMIN	Consensus-based Standards for the Selection of Health Measurement Instruments
CTT	Classical Test Theory
DIF	Differential Item Functioning
EFA	Exploratory Factor Analysis
GRADE	Grading of Recommendations Assessment, Development, and Evaluation
IRT	Item-Response Theory
LoA	Limits of Agreement
MEDLINE	Medical Literature Analysis and Retrieval System Online
MIC	Minimal Important Change
PBSC	Psychology and Behavioral Sciences Collection
PCA	Principal Component Analysis
PMC	PubMed Central
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM	Patient-Reported Outcome Measure
QoE	Quality of Evidence
RMSEA	Root Mean Square Error of Approximation
ROC	Receiver Operating Characteristic
SAMHSA	Substance Abuse and Mental Health Services Administration
SD	Standard Deviation
SDC	Smallest Detectable Change
SEM	Standard Error of Measurement
SMI	Serious Mental Illness
WHO	World Health Organization

Patient-reported outcome measures abbreviations	Full-term
BRQ	Bipolar Recovery Questionnaire
CHOICE	Choice of Outcome for Cognitive Behavioral Therapy in Psychosis
CRM	Consumer Recovery Measure
HAO	Hope, Agency and Opportunity
I.ROC	Individual Recovery Outcomes Counter
IMRS-C	Illness Management and Recovery Scale - Client version
MARS	Maryland Assessment of Recovery Scale
MHRM	Mental Health Recovery Measure
MHRS	Mental Health Recovery Star
MVML	MyVoiceMyLife
POP-RS	Peer Outcome Protocol - Recovery Subscale
PRI	Psychosis Recovery Inventory
QPR	Questionnaire about the Process of Recovery
RAS	Recovery Assessment Scale
RAS-DS	Recovery Assessment Scale - Domains and Stages
RI	Recovery Index
RMQ	Recovery Markers Questionnaire
RPI	Recovery Process Inventory
SAMHSA-RIC	Substance Abuse and Mental Health Services Administration - Recovery Inventory for Chinese
SISR	Self-Identified Stage of Recovery
SRS	Stages of Recovery Scale
STORI	Stages of Recovery Instrument
SubRAS	Subjective Recovery Assessment Scale
WBDRI	WestBridge Dual Recovery Inventory

Appendix B. Flow Diagram of Study Searches.



Note. Adapted from PRISMA 2020 flow diagram model (Page et al., 2021). PROM = Patient-Reported Outcome Measure.

Appendix C. Characteristics of Included PROMs.

PROM version	Reference	PROM description	Population of development	Format	Original language	PROM translations included in review	Conceptual basis of items	Users' involvement in development process
Recovery Assessment Scale (RAS)								
41 items	Corrigan et al. (1999)	Outcome measure for program evaluation	SMI	5-point Likert scale	English	Italian, French	Narrative analysis of four consumers' recovery stories; experts' agreement	Yes
24 items	Corrigan et al. (2004)	Psychometrically shortened version of RAS-41	SMI	5-point Likert scale	English	Japanese, Hebrew, French, Portuguese, Chinese, German, Cantonese, Spanish, Norwegian, Turkish, Dutch	–	–
Questionnaire about the Process of Recovery (QPR)								
22 items	Neil et al. (2009)	Collaboratively designed outcome measure of PR	Psychotic disorders	5-point Likert scale	English	Mandarin Chinese, Swedish, Japanese	Pitt et al. (2007) model of PR	Yes
15 items	Law et al. (2014)	Psychometrically shortened version of QPR-22	Schizophrenia	5-point Likert scale	English	Dutch, German, Spanish	–	–
Mental Health Recovery Measure (MHRM)								
30 items	Bullock & Young (2003)	Outcome measure of PR	SMI	5-point Likert scale	English	Chinese, Dutch, Portuguese	Young and Ensing (1999) model of PR	–
10 items	Armstrong et al. (2014)	Psychometrically shortened version of MHRM-30	SSD	5-point Likert scale	English	Turkish	–	–
STages Of Recovery Instrument (STORI)								
50 items	Andresen et al. (2006)	Theory driven outcome measure of processes and stages of PR	SMI	6-point Likert scale	English	Spanish, Hindi	Andresen et al. (2003) model of PR; literature review; experts' agreement	Yes
30 items	Andresen et al. (2013)	Psychometrically shortened version of STORI-50	SMI	6-point Likert scale	English	Chinese	–	–
Illness Management and Recovery Scale -Client version (IMRS-C)								
15 items	Gingerich and Mueser (2005)	Subscale of a larger inventory designed to assess outcomes of the Illness Management and Recovery (IMR) program.	SMI	5-point scales that vary across items	English	Hebrew, Dutch, Turkish	Items generated by IMR practitioners and consumers to reflect IMR outcomes	Yes
Peer Outcome Protocol – Recovery Scale (POP-RS)								
13 items	Campbell et al., (2004)	Subscale of the Peer Outcome Protocol, an inventory to assess outcomes of interest to people in recovery.	SMI	4-point Likert scale	English	Chinese	Literature review; focus group	Yes

(*continued on next page*)

(*continued*)

PROM version	Reference	PROM description	Population of development	Format	Original language	PROM translations included in review	Conceptual basis of items	Users' involvement in development process
Recovery Assessment Scale – Domains and Stages (RAS- DS)								
38 items	Hancock et al. (2015)	Outcome measure of PR, based on the RAS with additional items to cover later stages of PR	SMI	4-point Likert scale	English	–	RAS-24 items; focus groups	Yes
20 items	De Silva et al. (2021)	Psychometrically shortened version of RAS-DS-38	SMI	4-point Likert scale	English	–	–	–
Individual Recovery Outcomes Counter (I. ROC)								
12 items	Monger et al. (2012)	Outcome measure of PR designed for routine use	SMI	6-point visual analogue scale	English	Dutch	Experts' agreement; national agencies guidelines; existing tools	Yes
Subjective Recovery Assessment Scale (SubRAS)								
17 items	Yildiz et al. (2017)	Outcome measure of PR specifically designed for the Turkish cultural context.	SSD	5-point Likert scale	Turkish	–	Items from RAS, STORI & QPR; focus group with consumers and families	Yes
Psychosis Recovery Inventory (PRI)								
25 items	Chen, Tam, Wong, Law, & Chiu, 2005	Outcome measure of aspects related to the recovery of first-episode psychosis	Schizophrenia	6-point Likert scale	Chinese	English	Qualitative interviews with patients; experts' agreement	Yes
Recovery Process Inventory (RPI)								
22 items	Jerrell et al. (2006)	Outcome measure of PR to develop mental health system recovery orientation	SMI	5-point Likert scale	English	German	South Carolina Department of Mental Health operational definition of PR	Yes
Maryland Assessment of Recovery Scale (MARS)								
25 items	Drapalski et al. (2012)	Theory driven outcome measure of PR designed for both research and clinical purpose	SMI	5-point Likert scale	English	–	Experts' agreement; SAMHSA definition of PR	Yes
Mental Health Recovery Star (MHRS)								
10 items	Mckeith et al. (2008)	PR outcome measure designed to assess and summarize change, plan the actions needed to progress across domains	SMI	10-point scales that vary across items	English	–	–	Yes
Recovery Index (RI)								

(*continued on next page*)

(*continued*)

PROM version	Reference	PROM description	Population of development	Format	Original language	PROM translations included in review	Conceptual basis of items	Users' involvement in development process
21 items	IsHak et al. (2017)	IRT-made index composed of items from a social functioning scale (WSAS, Mundt et al., 2002) and a quality-of-life scale (Q-LES-Q, Endicott et al., 1993)	MDD	–	English	–	Items derived from two existing scales	–
WestBridge Dual Recovery Inventory (WBDRI)								
14 items	Noel et al. (2016)	PR outcome measure specifically designed for dual diagnosis program evaluation	Dual diagnosis of SMI and SUD	5-point Likert scale	English	–	Program's participants, staff and leaders generated items to capture program outcomes	Yes
SAMHSA-Recovery Inventory for Chinese (SAMHSA-RIC)								
111 items	Chiu et al. (2010)	Composite PR outcome scale made from 12 measures that assess recovery-related components	SSD	–	Cantonese	–	SAMHSA definition of PR; items derived from existing scales	–
41 items	Chiu et al. (2014)	Psychometrically IRT-reduced version of the SAMHSA-RIC-111	SSD	–	Cantonese	–	–	–
72 items	Chiu et al. (2020)	Psychometrically CFA-reduced version of the SAMHSA-RIC-111	SSD	–	Cantonese	–	–	–
Bipolar Recovery Questionnaire (BRQ)								
36 items	Jones et al. (2013)	PR outcome measure specifically designed for bipolar disorders	BD	Visual analogue scale	English	–	Literature review; qualitative study; experts' agreement	Yes
My Voice My Life (MVML)								
17 items	Gordon et al. (2013)	Consumer-led PR outcome measure	SMI	5-point Likert scale	English	–	Gordon et al. (2004) identified domains of PR; experts' agreement	Yes
Recovery Markers Questionnaire (RMQ)								
24 items	Ridgway & Press (2004)	Recovery subscale of the Recovery Enhancing Environment scale (REE/DREEM), a set of outcomes designed to be used to assess and promote organizational change towards recovery.	SMI	5-point Likert scale	English	Spanish	Consumers accounts of recovery; review of practices	Yes
CHOIce of outcome for CBT in psychosEs (CHOICE)								
24 items	Greenwood et al. (2010)	PR measure designed to assess key outcomes of CBTp programs	Psychotic disorders	10-point Osgood scale	English	–	Focus groups to identify key outcomes of the program; thematic and content analysis	Yes
11 items	Webb et al. (2021)	Psychometrically shortened version of CHOICE-24	SMI	10-point Osgood scale	English	–	–	–

(*continued on next page*)

(*continued*)

PROM version	Reference	PROM description	Population of development	Format	Original language	PROM translations included in review	Conceptual basis of items	Users' involvement in development process
Self-Identified Stage of Recovery – part A (SISR-A)								
1 item	Andresen et al. (2006)	Theory driven brief measure of PR stages	SMI	Nominal scale with 5 possible answers	English	Japanese	Andresen et al. (2003) model of PR	–
Self-Identified Stage of Recovery – part B (SISR-B)								
4 items	Andresen et al. (2006)	Theory driven brief measure of PR processes	SMI	6-point Likert scale	English	Japanese	Andresen et al. (2003) model of PR	–
Stages of Recovery Scale (SRS)								
45 items	Song & Hsu (2011)	Scale designed to assess the components of process and outcomes of PR, and to determine the person's stage of recovery, in order to promote recovery-oriented services	SMI	4-point Likert scale	Chinese	–	Experts' agreement; existing scales (STORI, RAS, MHRM)	Yes
Service-user Recovery Evaluation scale (SeRvE)								
40 items	Barber et al. (2012)	PR outcome scale with specific focus on spiritual well-being	SMI	5-point Likert scale	English	–	Literature search; existing scales; focus groups	Yes
15 items	Barber et al. (2017)	Shortened version of the SeRvE-40, with qualitative feedback from users	SMI	5-point Likert scale	English	–	–	Yes
Consumer Recovery Measure (CRM)								
15 items	Lusczakoski et al. (2016)	Brief PR outcome scale developed by a participatory work group for use in the holistic recovery evaluation system used internally by the Mental Health Center of Denver (MHCD), U.S.A.	SMI	4-point Likert scale	English	–	MHCD definition of recovery, focus groups	Yes
Hope, Agency, Opportunity (HAO)								
4 items	Newman-Taylor et al. (2017)	Very brief measure of recovery, combining patient reported outcomes and experience of services	Mental illness	5-point Likert scale	English	–	Perkins & Repper, 2003 model of mental health practice + literature search	Yes

Note. PROM = Patient-Reported Outcome Measure; CBTp = Cognitive Behavior Therapy for psychosis, CFA = Confirmatory Factor Analysis, IRT = Item-Response Theory, PR = Personal Recovery, SMI = Serious Mental Illness, SSD = Schizophrenia Spectrum Disorder, SUD = Substance Use Disorder, MDD = Major Depressive Disorder, SAMHSA = Substance Abuse and Mental Health Services Administration, IRT = Item Response Theory. Conceptual basis of items was extracted either from the PROM authors article or found in Campbell-Orde et al., (2005). "User's involvement in the development process" column reports a Yes if users were actually involved in the PROM development process (e.g., interviews, focus groups, pilot studies). Dashes indicate missing data.

Appendix D. Mapping of PROMs Items onto the CHIME Framework.

PROM version	Connectedness		Hope		Identity		Meaning in life		Empowerment		Non-mapping items		Agreement rate
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	%
RAS													
41 items	4	9,76	6	14,63	4	9,76	5	12,20	17	41,46	5	12,20	80.5
24 items	4	16,67	5	20,83	3	12,50	4	16,67	6	25,00	2	8,33	79.2
QPR													
22 items	6	27,27	2	9,09	3	13,64	6	27,27	5	22,73	0	0	63.7
15 items	2	13,33	2	13,33	3	20,00	5	33,33	3	20,00	0	0	66.6
MHRM													
30 items	3	10,00	3	10,00	6	20,00	8	26,67	5	16,67	5	16,67	66.6
10 items	0	0	0	0	4	40,00	3	30,00	0	0	3	30,00	70.0
STORI													
50 items	2	4,00	6	12,00	9	18,00	10	20,00	13	26,00	10	20,00	60.0
IMRS-Client													
15 items	2	13,33	2	13,33	4	26,67	7	46,67	2	13,33	2	13,33	73.3
POP-RS													
13 items	0	0	3	23,08	0	0	0	0	4	30,77	6	46,15	53.8
RAS-DS													
38 items	7	18,42	5	13,16	2	5,26	8	21,05	14	36,84	2	5,26	81.6
SubRAS													
17 items	6	35,29	2	11,76	1	5,88	5	29,41	3	17,65	0	0	82.3
PRI													
25 items	0	0	0	0	0	0	0	0	2	8,00	23	92,00	76.0
RPI													
22 items	6	27,27	4	18,18	2	9,09	4	18,18	3	13,64	3	13,64	81.2
MARS													
25 items	2	8,00	6	24,00	4	16,00	1	4,00	11	44,00	1	4,00	72.0
MHRS													
10 items	2	20,00	1	10,00	1	10,00	1	10,00	5	50,00	0	0	80.0
BRQ													
36 items	1	2,78	0	0	2	5,56	9	25,00	12	33,33	12	33,33	69.4
MVML													
65 items	13	20,00	1	1,54	8	12,31	22	33,85	20	30,77	1	1,54	66.1
RMQ													
24 items	4	16,67	2	8,33	1	4,17	10	41,67	3	12,50	4	16,67	79.2
CHOICE													
12 items	1	8,33	1	8,33	1	8,33	4	33,33	4	33,33	1	8,33	91.7
SISR-A&B													
9 items	0	0	0	0	1	11,11	1	11,11	1	11,11	6	66,67	88.9
SeRvE													
40 items	9	23,08	3	7,69	4	10,26	14	35,90	2	5,13	7	17,95	87.5
15 items	2	13.33	4	26.67	2	13.33	5	33.33	1	6.67	1	6.67	100
HAO													
4 items	1	25.00	1	25.00	0	0	1	25.00	1	25.00	0	0	100
CRM													
15 items	4	28,57	3	21,43	0	0	3	21,43	3	21,43	1	7,14	86.7

Note. PROMs items were reviewed independently by SF & KMV and attributed to a CHIME dimension when possible. Number (n) and percentage (%) of items mapped onto each dimension are displayed. Non-mapping items refers to the items that could not be attributed to any dimension. The STORI-30, RAS-SF-20, I.ROC-12, RI-21, WBDRI-14, SAMHSA-RIC-72 & 41, and SRS-45 were excluded from this analysis as we could not access the scales. PROM = Patient-Reported Outcome Measure; RAS = Recovery Assessment Scale; QPR = Questionnaire about the Process of Recovery; MHRM = Mental Health Recovery Measure; STORI = STages Of Recovery Instrument; IMRS-C = Illness Management and Recovery Scale -Client version; POP-RS = Peer Outcome Protocol – Recovery Scale; RAS-DS = Recovery Assessment Scale – Domains and Stages; I.ROC = Individual Recovery Outcomes Counter; SubRAS = Subjective Recovery Assessment Scale; PRI = Psychosis Recovery Inventory; RPI = Recovery Process Inventory; MARS = Maryland Assessment of Recovery Scale; MHRS = Mental Health Recovery Star; RI = Recovery Index; WBDRI = WestBridge Dual Recovery Inventory; SAMHSA-RIC = SAMHSA-Recovery Inventory for Chinese; BRQ = Bipolar Recovery Questionnaire; MVML = My Voice My Life; RMQ = Recovery Markers Questionnaire; CHOICE = CHOIce of outcome for CBT in psychoses; SISR-A = Self-Identified Stage of Recovery – part A; SISR-B = Self-Identified Stage of Recovery – part B; SRS = Stages of Recovery Scale; SeRvE = Service-user Recovery Evaluation scale; HAO = Hope Agency Opportunity; CRM = Consumer Recovery Measure.

Appendix E. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cpr.2024.102459.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. https://www.testingstandards.net.

Andresen, R., Caputi, P., & Oades, L. G. (2006). *Stages of recovery instrument : Development of a measure of recovery from serious mental illness* (p. 11).

Andresen, R., Oades, L., & Caputi, P. (2003). The experience of recovery from schizophrenia : Towards an empirically validated stage model. *The Australian and New Zealand Journal of Psychiatry, 37*(5), 586–594. https://doi.org/10.1046/j.1440-1614.2003.01234.x

Andresen, R., Oades, L. G., & Caputi, P. (2011). *Psychological recovery : Beyond mental illness* (1re éd). Wiley. https://doi.org/10.1002/9781119975182

Anthony, W. A. (1993). Recovery from mental illness : The guiding vision of the mental health service system in the 1990s. *Psychosocial Rehabilitation Journal, 16*(4), 11–23. https://doi.org/10.1037/h0095655

Anthony, W. A. (2000). A recovery-oriented service system : Setting some system level standards. *Psychiatric Rehabilitation Journal, 24*(2), 159–168. https://doi.org/10.1037/h0095104

Bachner-Melman, R., Lev-Ari, L., Zohar, A. H., & Linketsky, M. (2021). The eating disorders recovery questionnaire : Psychometric properties and validity. *Eating and Weight Disorders: EWD*. https://doi.org/10.1007/s40519-021-01139-y

Barber, J. M., Parkes, M., Parsons, H., & Cook, C. C. H. (2012). Importance of spiritual well-being in assessment of recovery : The service-user recovery evaluation (SeRvE) scale. *The Psychiatrist, 36*(12), 444–450. https://doi.org/10.1192/pb.bp.111.037838

Bassman, R. (2000). Agents, not objects : Our fights to be. *Journal of Clinical Psychology, 56*(11), 1395–1411. https://doi.org/10.1002/1097-4679(200011)56:11<1395::AID-JCLP3>3.0.CO;2-W

Beelen, L. M., van Dishoeck, A.-M., Tsangaris, E., Coriddi, M., Dayan, J. H., Pusic, A. L., … Vasilic, D. (2021). Patient-reported outcome measures in lymphedema : A

systematic review and COSMIN analysis. *Annals of Surgical Oncology, 28*(3), 1656–1668. https://doi.org/10.1245/s10434-020-09346-0

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research : A primer. *Frontiers in Public Health, 6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.

Burgess, P., Pirkis, J., Coombs, T., & Rosen, A. (2011). Assessing the value of existing recovery measures for routine use in Australian mental health services. *The Australian and New Zealand Journal of Psychiatry, 45*(4), 267–280. https://doi.org/10.3109/00048674.2010.549996

Campbell-Orde, T., Chamberlin, J., Carpenter, J., & Leff, H. S. (2005). *Measuring the Promise : A Compendium of Recovery Measures* (Volume II) (The Evaluation Center @ HSRI, Éd.) https://www.hsri.org//publication/measuring-the-promise-a-compendium-of-recovery-measures-volume-ii.

Carlton, J., Powell, P. A., & Project HERCULES Carer Group. (2022). Measuring carer quality of life in Duchenne muscular dystrophy : A systematic review of the reliability and validity of self-report instruments using COSMIN. *Health and Quality of Life Outcomes, 20*(1), 57. https://doi.org/10.1186/s12955-022-01964-4

Cavelti, M., Kvrgic, S., Beck, E.-M., Kossowsky, J., & Vauth, R. (2012). Assessing recovery from schizophrenia as an individual process. A review of self-report instruments. *European Psychiatry, 27*(1), 19–32. https://doi.org/10.1016/j.eurpsy.2011.01.007

Chen, E. Y. H., Tam, D. K. P., Wong, J. W. S., Law, C. W., & Chiu, C. P. Y. (2005). Self-administered instrument to measure the patient's experience of recovery after first-episode psychosis : Development and validation of the psychosis recovery inventory. *Australian & New Zealand Journal of Psychiatry, 39*(6), 493–499. https://doi.org/10.1080/j.1440-1614.2005.01609.x

Chien, W. T., & Chan, Z. C. Y. (2013). Chinese translation and validation of the questionnaire on the process of recovery in schizophrenia and other psychotic disorders. *Research in Nursing & Health, 36*(4), 400–411. https://doi.org/10.1002/nur.21549

Chirio-Espitalier, M., Schreck, B., Duval, M., Hardouin, J.-B., Moret, L., & Bronnec, M. G. (2022). Exploring the personal recovery construct in bipolar disorders : Definition, usage and measurement. A systematic review. *Frontiers in Psychiatry, 13*. https://www.frontiersin.org/articles/10.3389/fpsyt.2022.876761.

Chiu, M. Y. L., Ho, W. W. N., Lo, W. T. L., & Yiu, M. G. C. (2010). Operationalization of the SAMHSA model of recovery : A quality of life perspective. *Quality of Life Research, 13*.

Corrigan, P. W., Salzer, M., Ralph, R. O., Sangster, Y., & Keck, L. (2004). Examining the factor structure of the recovery assessment scale. *Schizophrenia Bulletin, 30*(4), 1035–1041. https://doi.org/10.1093/oxfordjournals.schbul.a007118

Crowley, K. (1997). *Implementing the concept of recovery, final report : Blue ribbon commission on mental health.* Madison: Wisconsin Department of Health Services.

Davidson, L., O'Connell, M. J., Tondora, J., Lawless, M., & Evans, A. C. (2005). Recovery in serious mental illness : A new wine or just a new bottle? *Professional Psychology: Research and Practice, 36*(5), 480–487. https://doi.org/10.1037/0735-7028.36.5.480

Drapalski, A. L., Medoff, D., Unick, G. J., Velligan, D. I., Dixon, L. B., & Bellack, A. S. (2012). Assessing recovery of people with serious mental illness : Development of a new scale. *Psychiatric Services, 63*(1), 48–53. https://doi.org/10.1176/appi.ps.201100109

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Fisher, D. B. (1994). Health care reform based on an empowerment model of recovery by people with psychiatric disabilities. *Hospital & Community Psychiatry, 45*(9), 913–915. https://doi.org/10.1176/ps.45.9.913

Gana, K., Boudouda, N. E., Ben Youssef, S., Calcagni, N., & Broc, G. (2021). Adaptation transculturelle de tests et échelles de mesure psychologiques : Guide pratique basé sur les Recommandations de la Commission Internationale des Tests et les Standards de pratique du testing de l'APA. *Pratiques Psychologiques, 27*(3), 223–240. https://doi.org/10.1016/j.prps.2021.02.001

Giffort, D., Schmook, A., Woody, C., Vollendorf, C., & Gervain, M. (1995). *Construction of a scale to measure consumer recovery*. Springfield, IL: Illinois Office of Mental Health.

Gilbody, S. M., House, A. O., & Sheldon, T. A. (2002). Psychiatrists in the UK do not use outcomes measures : National survey. *The British Journal of Psychiatry, 180*(2), 101–103. https://doi.org/10.1192/bjp.180.2.101

Gingerich, S., & Mueser, K. T. (2005). Illness management and recovery. In *Evidence-based mental health practice : A textbook* (pp. 395–424). W. W. Norton & Company.

Goh, Y. S. S., Ow Yong, J. Q. Y., & Li, A. Z. (2023). Effectiveness of illness management and recovery program on people with severe mental illnesses : A systematic review and meta-analysis. *Frontiers. Psychiatry, 14*. https://www.frontiersin.org/articles/10.3389/fpsyt.2023.1162288.

Gordon, S. E., Ellis, P. M., Siegert, R. J., & Walkey, F. H. (2012). Development of a self-assessed consumer recovery outcome measure : My voice, my life. *Administration and Policy in Mental Health, 40*(3), 199–210. https://doi.org/10.1007/s10488-012-0417-9

Greenwood, K. E., Sweeney, A., Williams, S., Garety, P., Kuipers, E., Scott, J., & Peters, E. (2010). CHoice of Outcome In Cbt for psychosEs (CHOICE) : The Development of a New Service User–Led Outcome Measure of CBT for Psychosis. *Schizophrenia Bulletin, 36*(1), 126–135. https://doi.org/10.1093/schbul/sbp117

Gruijters, S. L. K. (2020). Using principal component analysis to validate psychological scales : Bad statistical habits we should have broken yesterday II [preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/qbe6k

Hancock, N., Bundy, A., Honey, A., Helich, S., & Tamsett, S. (2013). Measuring the later stages of the recovery journey : Insights gained from clubhouse members. *Community Mental Health Journal, 49*(3), 323–330. https://doi.org/10.1007/s10597-012-9533-y

Hancock, N., Scanlan, J. N., Honey, A., Bundy, A. C., & O'Shea, K. (2015). Recovery Assessment Scale—Domains and Stages (RAS-DS) : Its feasibility and outcome measurement capacity. *The Australian and New Zealand Journal of Psychiatry, 49*(7), 624–633. https://doi.org/10.1177/0004867414564084

Harrison, J. L., Brownlow, C. L., Ireland, M. J., & Piovesana, A. M. (2022). Empathy measurement in autistic and nonautistic adults : A COSMIN systematic literature review. *Assessment, 29*(2), 332–350. https://doi.org/10.1177/1073191120964564

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis : Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Hunter, R. H., & Marsh, D. T. (1994). Mining giftedness : A challenge for psychologists. In *New directions in the psychological treatment of serious mental illness* (pp. 99–122). Praeger Publishers/Greenwood Publishing Group.

International Test Commission. (2017). *ITC Guidelines for Translating and Adapting Tests* (2nd ed.) https://www.intestcom.org.

IsHak, W. W., Bonifay, W., Collison, K., Reid, M., Youssef, H., Parisi, T., … Cai, L. (2017). The recovery index : A novel approach to measuring recovery and predicting remission in major depressive disorder. *Journal of Affective Disorders, 208*, 369–374. https://doi.org/10.1016/j.jad.2016.08.081

Jensen-Doss, A., Smith, A. M., Becker-Haimes, E. M., Mora Ringle, V., Walsh, L. M., Nanda, M., … 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 and Mental Health Services Research, 45*(3), 392–403. https://doi.org/10.1007/s10488-017-0833-y

Koğar, H. (2020). Development of a short form : Methods, examinations, and recommendations. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 11*(3), 302–310. https://doi.org/10.21031/epod.739548

Krägeloh, C. U., Czuba, K. J., Billington, D. R., Kersten, P., & Siegert, R. J. (2015). Using feedback from patient-reported outcome measures in mental health services : A scoping study and typology. *Psychiatric Services, 66*(3), 224–241. https://doi.org/10.1176/appi.ps.201400141

Law, H., Morrison, A., Byrne, R., & Hodson, E. (2012). Recovery from psychosis : A user informed review of self-report instruments for measuring recovery. *Journal of Mental Health, 21*(2), 192–207. https://doi.org/10.3109/09638237.2012.670885

Law, H., Neil, S. T., Dunn, G., & Morrison, A. P. (2014). Psychometric properties of the questionnaire about the process of recovery (QPR). *Schizophrenia Research, 156*(2–3), 184–189. https://doi.org/10.1016/j.schres.2014.04.011

Leamy, M., Bird, V., Boutillier, C. L., Williams, J., & Slade, M. (2011). Conceptual framework for personal recovery in mental health : Systematic review and narrative synthesis. *British Journal of Psychiatry, 199*(6), 445–452. https://doi.org/10.1192/bjp.bp.110.083733

Leendertse, J. C. P., Wierdsma, A. I., van den Berg, D., Ruissen, A. M., Slade, M., Castelein, S., & Mulder, C. L. (2021). Personal recovery in people with a psychotic disorder : A systematic review and Meta-analysis of associated factors. *Frontiers in Psychiatry, 12*, Article 622628. https://doi.org/10.3389/fpsyt.2021.622628

Lim, M. S. M., Byrne, A., Shieh, J., Hồ, Q. T., & Mason, O. (2017). Validation of a Vietnamese Mental Health Recovery Scale for Vietnamese refugees. *Journal of Psychosocial Rehabilitation and Mental Health, 4*(2), 135–145. https://doi.org/10.1007/s40737-017-0086-9

Llewellyn-Beardsley, J., Rennick-Egglestone, S., Callard, F., Pollock, K., Slade, M., & Edgley, A. (2023). "Nothing's changed, baby" : How the mental health narratives of people with multiple and complex needs disrupt the recovery framework. *SSM - Mental Health, 100221*. https://doi.org/10.1016/j.ssmmh.2023.100221

Lusczakoski, K. K., Olmos-Gallo, P. A., Milnor, W., & McKinney, C. J. (2016). Measuring mental health recovery : An application of Rasch modeling to the consumer recovery measure. *The Journal of Behavioral Health Services & Research, 43*(1), 104–115. https://doi.org/10.1007/s11414-014-9411-1

Meinck, F., Neelakantan, L., Steele, B., Jochim, J., Davies, L. M., Boyes, M., … Dunne, M. (2022). Measuring violence against children : A COSMIN systematic review of the psychometric properties of child and adolescent self-report measures. *Trauma, Violence & Abuse, 15248380221082152*. https://doi.org/10.1177/15248380221082152

Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2017). COSMIN risk of Bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research: an International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 27*(5), 1171–1179. https://doi.org/10.1007/s11136-017-1765-4

Mokkink, L. B., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) : User manual. https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., … de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63*(7), 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Monger, B., Hardie, S. M., Ion, R., Cumming, J., & Henderson, N. (2013). The individual recovery outcomes counter : Preliminary validation of a personal recovery measure. *Psychiatrist, 37*(7), 221–227. Scopus https://doi.org/10.1192/pb.bp.112.041889.

National Institute of Mental Health. (2024). Mental Illness. https://www.nimh.nih.gov/health/statistics/mental-illness.

Neil, S., Kilbride, M., Pitt, L., Nothard, S., Welford, M., Sellwood, W., & Morrison, A. (2009). The questionnaire about the process of recovery (QPR) : A measurement tool

developed in collaboration with service users. *Psychosis, 1*, 145–155. https://doi.org/10.1080/17522430902913450
Newman-Taylor, K., Garner, C., Vernon-Wilson, E., Paas, K. H. W., Herbert, L., & Au-Yeung, S. K. (2017). Psychometric evaluation of the hope, agency and opportunity (HAO); a brief measure of mental health recovery. *Journal of Mental Health (Abingdon, England), 26*(6), 562–568. https://doi.org/10.1080/09638237.2017.1385746
Newman-Taylor, K., Maguire, T., & Bowen, A. (2019). Why are we not measuring what matters in mental health in the UK? The case for routine use of recovery outcome measures. *Perspectives in Public Health, 139*(4), 181–183. https://doi.org/10.1177/1757913919851414
Noel, V., Woods, M., Routhier, J., & Drake, R. (2016). Planning treatment and assessing recovery in participants with dual diagnosis : Preliminary evaluation of a new clinical tool. *Journal of Dual Diagnosis, 12*(1), 55–62. https://doi.org/10.1080/15504263.2016.1146555
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement : An updated guideline for reporting systematic reviews. *BMJ, n71*. https://doi.org/10.1136/bmj.n71
Penas, P., Iraurgi, I., Concepción Moreno, M., & Uriarte, J. J. (2019). How is evaluated mental health recovery? : A systematic review. *Actas Espanolas de Psiquiatria*, *47*(1), 23-32. Scopus.
Pitt, L., Kilbride, M., Nothard, S., Welford, M., & Morrison, A. P. (2007). Researching recovery from psychosis : A user-led project. *Psychiatric Bulletin, 31*(2), 55–60. https://doi.org/10.1192/pb.bp.105.008532
Prinsen, C. A. C., Mokkink, 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
Rezai, M., Kolne, K., Bui, S., & Lindsay, S. (2020). Measures of workplace inclusion : A systematic review using the COSMIN methodology. *Journal of Occupational Rehabilitation, 30*(3), 420–454. https://doi.org/10.1007/s10926-020-09872-4
Roe, D., Slade, M., & Jones, N. (2022). The utility of patient-reported outcome measures in mental health. *World Psychiatry, 21*(1), 56–57. https://doi.org/10.1002/wps.20924
Rudd, B., Karatzias, T., Bradley, A., Fyvie, C., & Hardie, S. (2020). Personally meaningful recovery in people with psychological trauma : Initial validity and reliability of the individual recovery outcomes counter (I.ROC). *International Journal of Mental Health Nursing, 29*(3), 387–398. https://doi.org/10.1111/inm.12671
Saputra, A., Kusumawardhani, A. A. A. A., Elvira, S. D., & Wiguna, T. (2022). An item development, content validity, and feasibility study towards the Indonesian recovery scale for patients with schizophrenia. *Heliyon, 8*(11), Article e11826. https://doi.org/10.1016/j.heliyon.2022.e11826
Scheyett, A., DeLuca, J., & Morgan, C. (2013). Recovery in severe mental illnesses : A literature review of recovery measures. *Social Work Research, 37*(3), 286–303. https://doi.org/10.1093/swr/svt018
Shanks, V., Williams, J., Leamy, M., Bird, V. J., Le Boutillier, C., & Slade, M. (2013). Measures of personal recovery : A systematic review. *Psychiatric Services, 64*(10), 974–980. https://doi.org/10.1176/appi.ps.005012012
Sklar, M., Groessl, E. J., O'Connell, M., Davidson, L., & Aarons, G. A. (2013). Instruments for measuring mental health recovery : A systematic review. *Clinical Psychology Review, 33*(8), 1082–1095. https://doi.org/10.1016/j.cpr.2013.08.002
Slade, M. (2010). Measuring recovery in mental health services. *The Israel Journal of Psychiatry and Related Sciences, 47*(3), 206–212.
Slade, M., Adams, N., & O'Hagan, M. (2012). Recovery : Past progress and future challenges. *International Review of Psychiatry, 24*(1), 1–4. https://doi.org/10.3109/09540261.2011.644847
Slade, M., Amering, M., Farkas, M., Hamilton, B., O'Hagan, M., Panther, G., Perkins, R., Shepherd, G., Tse, S., & Whitley, R. (2014). Uses and abuses of recovery : Implementing recovery-oriented practices in mental health systems. *World Psychiatry, 13*(1), 12–20. https://doi.org/10.1002/wps.20084
Slade, M., Leamy, M., Bacon, F., Janosik, M., Le Boutillier, C., Williams, J., & Bird, V. (2012). International differences in understanding recovery : Systematic review. *Epidemiology and Psychiatric Sciences, 21*(4), 353–364. Scopus https://doi.org/10.1017/S2045796012000133.
Song, L. (2017). Predictors of personal recovery for persons with psychiatric disabilities : An examination of the Unity model of recovery. *Psychiatry Research, 250*, 185–192. https://doi.org/10.1016/j.psychres.2017.01.088
Song, L.-Y., & Shih, C.-Y. (2009). Factors, process and outcomes of recovery from psychiatric disability : The Unity model. *International Journal of Social Psychiatry, 55*(4), 348–360. https://doi.org/10.1177/0020764008093805
Sportel, B. E., Aardema, H., Boonstra, N., Arends, J., Rudd, B., Metz, M. J., … Pijnenborg, G. H. M. (2023). Measuring recovery in participants with a schizophrenia spectrum disorder : Validation of the individual recovery outcomes counter (I.ROC). *BMC Psychiatry, 23*(1), 296. https://doi.org/10.1186/s12888-023-04763-3
Tenney, L. J. (2000). It has to be about choice. *Journal of Clinical Psychology, 56*(11), 1433–1445. https://doi.org/10.1002/1097-4679(200011)56:11<1433::AID-JCLP6>3.0.CO;2-5
Terwee, C. B., Jansma, E. P., Riphagen, I. I., & de Vet, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research, 18*(8), 1115–1123. https://doi.org/10.1007/s11136-009-9528-5
Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., … Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures : A Delphi study. *Quality of Life Research: an International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 27*(5), 1159–1170. https://doi.org/10.1007/s11136-018-1829-0
Thomas, E. C., Despeaux, K. E., Drapalski, A. L., & Bennett, M. (2018). Person-oriented recovery of individuals with serious mental illnesses : A review and Meta-analysis of longitudinal findings. *Psychiatric Services, 69*(3), 259–267. https://doi.org/10.1176/appi.ps.201700058
Davidson, L., Rowe, M., & Tondora, J. (2008). *A Practical Guide to Recovery-Oriented Practice: Tools for Transforming Mental Health Care*. Oxford University Press.
De Silva, K. T., Cockshaw, W. D., Rehm, I. C., & Hancock, N. (2021). A short form of the Recovery Assessment Scale-Domains and Stages : Development and validation among adults with anxiety disorders. *Clinical Psychology & Psychotherapy*. https://doi.org/10.1002/cpp.2563
de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine : A practical guide*. Cambridge University Press. https://doi.org/10.1017/CBO9780511996214
Williams, J., Leamy, M., Bird, V., Le Boutillier, C., Norton, S., Pesola, F., & Slade, M. (2015). Development and evaluation of the INSPIRE measure of staff support for personal recovery. *Social Psychiatry and Psychiatric Epidemiology, 50*(5), 777–786. https://doi.org/10.1007/s00127-014-0983-0
Wolstencroft, K., Oades, L., Caputi, P., & Andresen, R. (2010). Development of a structured interview schedule to assess stage of psychological recovery from enduring mental illness. *International Journal of Psychiatry in Clinical Practice, 14*(3), 182–189. https://doi.org/10.3109/13651501003668552
World Health Organization. (2021). *Guidance on community mental health services : Promoting person-centred and rights-based approaches*. In *Guidance and technical packages on community mental health services: Promoting person-centred and rights-based approaches*. World Health Organization. https://apps.who.int/iris/handle/10665/341648.
Yildiz, M., Erim, R., Soygur, H., Tural, U., Kiras, F., & Gules, E. (2017). Development and validation of the subjective recovery assessment scale for patients with schizophrenia. *Psychiatry and Clinical Psychopharmacology, 28*(2), 163–169. APA PsycInfo https://doi.org/10.1080/24750573.2017.1405579.
Young, S. L., & Ensing, D. S. (1999). Exploring recovery from the perspective of people with psychiatric disabilities. *Psychiatric Rehabilitation Journal, 22*(3), 219–231. https://doi.org/10.1037/h0095240