European Addiction Research

Research Article

Eur Addict Res
DOI: 10.1159/000541584

Received: March 15, 2024
Accepted: September 20, 2024
Published online: November 8, 2024

Psychometric Evaluation of the Dutch Version of the Substance Use Recovery Evaluator (SURE-NL)

Charlotte Migchels[a] Wim van den Brink[b] Amine Zerrouk[c]
Frieda I.A. Matthys[a] Clara De Ruysscher[c] Dries Debeer[d]
Wouter Vanderplasschen[c] Cleo Lina Crunelle[a]

[a]Department of Psychiatry, Vrije Universiteit Brussel (VUB), Universitair Ziekenhuis Brussel (UZ Brussel), Brussels, Belgium; [b]Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands; [c]Department of Special Needs Education, Ghent University (UGent), Ghent, Belgium; [d]Faculty of Psychology and Educational Sciences, Ghent University (UGent), Ghent, Belgium

Keywords
Patient-reported outcome measures · Substance use disorder · Recovery · Psychometric evaluation · Substance use recovery evaluator

Abstract
Introduction: Patient-reported outcome measures (PROMs) are an important source of information that allow for a patient-centered assessment. Outcome measurement of substance use disorder (SUD) treatment traditionally focuses on drug use and deficits in functioning, overlooking other aspects of the personal and dynamic process of recovery. The substance use recovery evaluator (SURE) is a PROM developed with service user input to monitor the recovery journey and assess treatment outcomes in people with SUD. The objective of this study was to examine the validity and reliability of the Dutch translation of the SURE, the SURE-NL, for measuring indicators of recovery in Dutch-speaking patients with SUD. ***Methods:*** The original SURE questionnaire was translated from English to Dutch using forward-backward translation. A total of *N* = 171 participants were recruited as part of a naturalistic multicenter study in inpatient (*N* = 149) and outpatient (*N* = 22) SUD treatment centers. We examined the factorial structure, reliability, and concurrent and discriminant validity of the SURE-NL. ***Results:*** The original 5-factor structure of the SURE showed acceptable fit for the SURE-NL, and internal consistencies of the subscales ranged from 0.61 to 0.76; internal consistency of the total score was 0.83. Concurrent validity was confirmed through positive correlations of the SURE-NL total and subscale scores with the WHOQoL-BREF subscales, but correlations varied depending on subscale and treatment setting, with higher correlations for the outpatient compared to the inpatient subgroup. Discriminant validity was confirmed through low, mostly non-significant correlations between the SURE-NL and the DASS-21. ***Conclusions:*** Although the SURE was originally designed for outpatient settings, our findings in a predominantly inpatient sample indicate that the SURE-NL is suitable for assessing personal recovery and recovery capital in Dutch-speaking Belgian patients with SUD. However, subscales should be used and interpreted with caution. Further research is needed with

karger@karger.com
www.karger.com/ear

Correspondence to:
Wim van den Brink, w.vandenbrink@amsterdamumc.nl

larger Dutch-speaking outpatient samples and the development of a tailored SURE for inpatient settings should be considered.

Introduction

People with substance use disorder (SUD) are at high risk of adverse consequences, including other mental and physical disorders, road-traffic accidents, violence, unemployment, poverty, and homelessness [1, 2]. Yet, outcome measures in SUD treatment traditionally focus on abstinence or the reduction of drug use, overlooking other important domains impacted by drug and/or alcohol use, such as general health, well-being, and relationships [3–5]. It has been recommended to routinely include these other domains of functioning as essential outcomes in SUD treatment studies [3, 4, 6]. Existing outcome measures, however, do not always target the problems experienced by patients with SUD and their loved ones, but rather reflect the perspectives of clinicians and researchers [7].

Recently, the use of patient-reported outcome measures (PROMs) has increased in all healthcare areas, emphasizing the significance of patients' own views on health and well-being [7–9]. PROMs are questionnaires that collect information on health outcomes, including symptoms, functional status, and quality of life, directly from patients [8]. PROMs are an important source of information that can be used for a wide range of purposes, from clinical practice and clinical trials to public health and policy evaluation [10]. Developing PROMs that represent and incorporate the patient's perspective and are psychometrically robust is a complex but necessary process, if we aim for a truly patient-centered assessment [7, 9, 10].

The emerging recovery paradigm in the SUD field illustrates a similar trend and focuses beyond abstinence, and is increasingly being integrated in SUD research, policy, and practice [11, 12]. Recovery is viewed as a deeply personal and dynamic process of change, encompassing several life domains [13]. It entails multiple interrelated aspects, including clinical recovery (e.g., reduced substance use, abstinence), functional recovery (e.g., housing, daily activities, employment), personal recovery (e.g., connectedness, hope, identity, meaning, empowerment), and social recovery (e.g., overcoming stigma, inclusive citizenship) [12, 14, 15]. From this perspective, abstinence is not the only possible pathway to recovery, shifting the emphasis toward building on available strengths, expanding recovery capital, and creating nurturing contexts in which persons can transit from active addiction to recovery [16, 17].

In 2016, Neale et al. [11] introduced the Substance Use Recovery Evaluator (SURE), a PROM developed with service user input to monitor the recovery journey and assess treatment outcomes in people with SUD. The SURE is a 21-item questionnaire with a total score and five relatively independent factors: "substance use," "self-care," "relationships," "material resources," and "outlook on life." To the best of our knowledge, the only validated translation of the SURE that is currently available is a German version with 20 (instead of 21) items and three (instead of five) factors: "psychological and physical well-being," "daily functioning," and "substance use" [18]. These differences underscore the importance of validating instruments before their application in a culturally and linguistically different population from the original context in which they were developed.

Recovery has gained increasing attention in SUD treatment in Belgium and the Netherlands [17]. A psychometrically sound translated version of the SURE could enable patient-centered routine assessment of recovery indicators in people with SUD within Dutch-speaking populations. Therefore, the objective of this study was to translate the SURE to Dutch (SURE-NL) and evaluate the psychometric properties of this translation in a sample of Dutch-speaking patients from various SUD treatment settings in Belgium. The psychometric evaluation included statistical analyses (a) to establish the underlying factor structure of the SURE-NL and (b) to assess the reliability (Cronbach's α) of the different subscales, and correlational analyses to (c) explore the concurrent validity of the SURE-NL using the World Health Organization Quality of Life questionnaire – brief (WHOQoL-BREF) and (d) the discriminant validity using the Depression, Anxiety, Stress Scale (DASS-21) [19, 20].

Methods

Translation

The original SURE questionnaire was translated from English to Dutch using forward-backward translation in accordance with international guidelines [21]. Forward translation was conducted by C.L.C. and W.v.d.B., native Dutch speakers and experts in the SUD field. Next, an independent back translation was performed by a native English speaker. Following each step, translations were discussed by an expert committee and adapted

accordingly. The preliminary translated questionnaire was used for pilot testing in five native Dutch-speaking patients who were in inpatient treatment for SUD. The pilot testing was reviewed by the Expert Committee and no modifications were deemed necessary, resulting in the final translated version of the questionnaire, the SURE-NL.

Measures

Demographic and Clinical Characteristics

Participants provided demographic information including age, sex, education level, and ethnicity. Self-completed clinical information included treatment history, participation in opioid agonist treatment (OAT), and main substance(s) used, with the option to select multiple substances if applicable.

Substance Use Recovery Evaluator

The SURE consists of 21 items that are completed using a 5-point Likert type response scale, but scored on a 3-point scale (1–3). The decision to use this 3-point rating scale was based on the low response probabilities of the two extreme response categories. The first 2 response options correspond to a score of 3, the third response option to a score of 2 and the final 2 response options to a score of 1. Response options for the first 3 questions are "never," "on 1 or 2 days," "on 3 or 4 days," "on 5 or 6 days," and "every day," while for the remaining questions response options are "all of the time," "most of the time," "a fair amount of the time," "a little of the time," and "none of the time." Total scores range from 21 to 63, with higher scores indicating more recovery strengths. Based on the factor analysis of the original SURE study, the 21 items are grouped in five subscales: "substance use" (6 items), "self-care" (5 items), "relationships" (4 items), "material resources" (3 items), and "outlook on life" (3 items) [11]. In a sample of current and former outpatient SUD service users in the UK, the SURE showed good construct validity and internal consistency was found to be acceptable to high for all subscales (Cronbach's α = 0.68 to 0.87) and high for the total score (Cronbach's α = 0.92) [11].

World Health Organization Quality of Life Questionnaire – Brief

The WHOQoL-BREF is a 26-item abbreviated version of the WHOQoL-100 assessment, assessing four domains of quality of life: "physical health" (7 items), "psychological health" (6 items), "social relationships" (3 items), and "environment" (8 items). All items are scored on a 5-point Likert type response scale (1–5). Response options, from lowest to highest score, are "very poor/very dissatisfied/not at all/never," "poor/dissatisfied/a little/seldom," "neither poor nor good/neither satisfied nor dissatisfied/a moderate amount/moderately/quite often," "good/satisfied/very much/mostly/very often," and "very good/very satisfied/an extreme amount/extremely/completely/always." Higher total and subscale scores indicate higher quality of life. The first 2 questions are scored separately, and 3 items are reverse-scored. In an international sample of healthcare users and people from the general population, the WHOQoL-BREF showed good construct and discriminative validity, and internal consistency was acceptable for all WHOQoL-BREF domains (Cronbach's α = 0.68 to 0.82) [19].

Depression, Anxiety, Stress Scale

The DASS-21 is a 21-item questionnaire, assessing self-reported symptoms of depression, anxiety, and stress. It consists of three subscales ("depression," "anxiety," and "stress") with 7 items each. The questions are rated on a scale from 0 (did not apply to me at all) to 3 (applied to me very much) [20]. Sum scores are calculated by adding up the scores on the individual items and multiplying them by a factor 2. In a Dutch population of SUD patients in residential detoxification, the DASS-21 total score was found to have high reliability (Cronbach's α = 0.92) [22].

Participants

A convenience sample of N = 171 participants was recruited as part of the Outcome Measurement and Evaluation as a Routine practice in alcohol and other drug services in Belgium (OMER-BE) study: a naturalistic multicenter study assessing SUD treatment services using patient-reported outcome and experience measures (PROMs and PREMs) [23]. Participants were included between July 2022 and September 2023 if they met the following inclusion criteria: (i) started treatment for SUD in the participating center less than 3 weeks before study inclusion; (ii) over 18 years old; and (iii) Dutch-speaking.

Participants were included from different inpatient (N = 149) and outpatient (N = 22) treatment modalities. Participants in inpatient treatment were recruited from three psychiatric centers and four therapeutic communities in Flanders. Therapeutic communities for SUD are "a drug-free environment in which people with addictive problems live together in an organized and structured way in order to promote change and make it possible for them to lead a drug-free life in the outside society" [24]. The five participating outpatient treatment centers offered individual treatment, with three centers also providing OAT.

Upon treatment entry, participants were asked by their treatment provider if they were interested in participating in the study. If they agreed to participate, the treatment provider contacted the researchers. Participants were given a tablet by the researcher on which they self-completed sociodemographic data, clinical information, and outcome measures, including the SURE-NL, the WHOQoL-BREF, and the DASS-21, using an online survey administered through LimeSurvey [25]. Participants filled out the questionnaires independently, but a researcher remained present throughout the assessment to respond to any queries from the participants. After completing all questionnaires, participants received a gift card with a value of EUR 10.

Statistical Analyses

All data analyses were conducted using IBM SPSS statistics version 29 and R statistical software (package lavaan) [26]. We opted for analyses based on classical test theories over those based on Item Response Theory (IRT) due to our limited sample size of $N = 171$, which may be insufficient for robust IRT modeling. Normality of data distribution was assessed. Demographics and clinical data were analyzed using descriptive statistics. Confirmatory factor analysis (CFA) was performed to test whether previously obtained factor structures of the original English SURE and the German translation of the SURE showed a good fit for the SURE-NL [11, 18]. Responses to the items were considered ordered categorical, and diagonally weighted least squares was used. Model fit was assessed using measures of both absolute and relative fit, i.e., the Root Mean Square Error of Approximation (RMSEA, values less than 0.08 indicate an acceptable fit), the scaled χ^2 test, the scaled version of the Comparative Fit Index (CFI, values of 0.90 and higher indicate an acceptable fit), and the Tucker-Lewis Index (TLI, values of 0.90 and higher indicate an acceptable fit) [27]. Reliability of the SURE-NL in terms of internal consistency was assessed by calculating Cronbach's α. Values between 0.70 and 0.95 indicate good reliability [28]. To examine concurrent validity, Spearman's correlations were used to assess the relationship of the SURE-NL total and subscale scores with the four subscales of the WHOQoL-BREF. Discriminant validity was assessed by examining Spearman's correlations of the SURE-NL total and subscale scores with the three DASS subscale scores. Participants' scores on the SURE-NL were calculated and a Mann-Whitney U test was used to assess whether there were differences in SURE-NL scores between participants in inpatient and outpatient treatment.

Table 1. Demographic and clinical characteristics of participant sample ($N = 171$)

Total sample ($N = 171$)	Mean	SD
Age	34.74	9.49
	N	%
Sex		
Male	142	83
Female	29	17
Country of birth		
Belgium	164	95.9
Other	7	4.1
Education level		
Primary education	40	23.4
Secondary education	105	61.4
Higher education	26	15.2
Type of treatment center		
Psychiatric inpatient center	69	40.4
Therapeutic community	80	46.8
Outpatient treatment	22	12.8
Opioid agonist therapy	28	16.4
Previous treatment for SUD	139	81.3
Main substance(s) of use[a,b]		
Cocaine	87	52.1
Alcohol	87	52.1
Cannabis	56	33.5
Amphetamines	42	25.2
Opioids	29	17.4
Ketamine	20	12
Benzodiazepines	19	11.4
Gamma-hydroxybutyrate (GHB)	13	7.8
Other	5	3

[a]Some participants reported more than one main substance used. [b]Data missing for 4 participants (total $N = 167$).

Results

Demographic and Clinical Characteristics

The majority of participants were male ($N = 142$, 83%), and the average age was 34.7 years (SD 9.5). For most, their highest level of education was secondary education ($N = 105$, 61.4%). Of the participants, 87.2% ($N = 149$) were in inpatient treatment and 12.8% ($N = 22$) in outpatient treatment. 81.3% ($N = 139$) had received previous treatment for SUD, and 16.4% received OAT at the time of assessment. Cocaine (52.1%) and alcohol (52.1%) were the most frequently used substances and 60.5% of the participants reported using more than one substance, with a median of 2 substances

Table 2. SURE-NL estimated factor score correlations

SURE-NL	Substance use	Self-care	Relationships	Material resources
Self-care	0.65			
Relationships	0.54	0.64		
Material resources	0.51	0.33	0.36	
Outlook on life	0.48	0.60	0.83	0.31

(range: 1–7). Table 1 provides an overview of the demographic and clinical characteristics of the participant sample.

Factor Structure

A CFA was performed on the full sample (N = 171). Both the original 5-factor structure as reported by Neale et al. [11] and the 3-factor structure of the German translation by Reichl et al. [18], as well as a 1-factor structure, were considered and their fit to our data was compared. Although Reichl et al. [18] excluded one item, we included all items in the models to make the fit of the models comparable. All items were considered ordered categorical items with three categories. The original 5-factor structure showed the best fit to the data with RMSEA = 0.071 (90% CI = [0.059–0.083]), TLI = 0.86 and CFI = 0.88, indicating acceptable fit. However, the chi-square test was significant ($\chi^2(179) = 331.39$, $p < 0.001$), indicating some degree of model misfit, but this test is known to be sensitive to sample size. In contrast, the 3-factor solution from the study on the German translation of the SURE did not show good fit (RMSEA = 0.086 (90% CI = [0.075–0.097]), TLI = 0.79, CFI = 0.81, $\chi^2(186) = 420.66$, $p < 0.001$). Finally, the 1-factor structure also did not show good fit (RMSEA = 0.093 (90% CI = [0.083–0.104]), TLI = 0.75, CFI = 0.78, $\chi^2(189) = 468.37$, $p < 0.001$). Thus, only the 5-factor model as proposed by Neale et al. [11] was confirmed in our sample by the CFA. Online supplementary material 1 (for all online suppl. material, see https://doi.org/10.1159/000541584) shows the CFA standardized loadings on the five factors.

Additionally, we conducted a CFA on the inpatient subsample (N = 149), and the results were largely consistent with those for the total sample. Again, the original 5-factor structure showed the best fit to the data, though the fit was slightly less robust than that observed in the total sample (RMSEA = 0.077 (90% CI = [0.064–0.089]), TLI = 0.82, CFI = 0.84, and $\chi^2(179) = 334.82$, $p < 0.001$). Considering the limited sample size, we did not perform a separate CFA for the outpatient subsample (N = 22).

The five factors showed positive, moderate to strong intercorrelations, ranging from 0.31 to 0.83 (Table 2), indicating that some factors are relatively independent, whereas others show considerable overlap in the aspects that are assessed. The observed subscales and total score correlations showed strong positive correlations between the SURE-NL subscales and the total score, ranging from 0.51 to 0.77 (Table 3). Based on these findings, we used both subscale scores and the total score in the rest of the analyses. Online supplementary material 2 shows the observed subscale and total score correlations for the inpatient subsample.

Reliability

The SURE-NL subscales "self-care" (5 items) and "outlook on life" (3 items) both had a Cronbach's α of 0.76, indicating good internal consistency (Table 3) [28]. The "relationships" subscale (4 items) had a Cronbach's α of 0.66, while the "substance use" (6 items) and the "material resources" (3 items) subscales had relatively low internal consistencies of Cronbach's α = 0.62 and Cronbach's α = 0.61, respectively. The 21-item total score of the SURE-NL showed good internal consistency (Cronbach's α = 0.83).

Concurrent Validity

Internal consistencies of the WHOQoL-BREF subscales calculated in our sample varied: "physical health" Cronbach's α = 0.70; "psychological health" Cronbach's α = 0.81; "social relationships" Cronbach's α = 0.56; and "environment" Cronbach's α = 0.82. For the total sample, the 21-item SURE-NL total score and subscale scores correlated positively with most WHOQoL-BREF subscale scores (Table 4). The SURE-NL subscales "relationships" and "outlook on life," as well as the SURE-NL total score, showed substantial correlations with the WHOQoL-BREF subscales, ranging from 0.28 to 0.58, while the correlations of the SURE-NL "substance use" and "material resources" subscales with the WHOQoL-BREF subscales were mostly small and non-significant. For

Table 3. SURE-NL observed subscale and total score correlations

SURE-NL	Substance use	Self-care	Relationships	Material resources	Outlook on life
Self-care	0.39				
Relationships	0.31	0.44			
Material resources	0.28	0.19	0.19		
Outlook on life	0.26	0.45	0.54	0.19	
Total score	0.72	0.77	0.68	0.51	0.68
α	0.62	0.76	0.66	0.61	0.76

Table 4. Spearman's rho correlation coefficients for the SURE-NL total and subscale scores and WHOQoL-BREF subscale scores

	SURE-NL					
	substance use	self-care	relationships	material resources	outlook on life	total score
WHOQOL-BREF						
Total sample (N = 171)						
Physical health	0.18*	0.37**	0.34**	0.05 (0.252)	0.49**	0.41**
Psychological health	0.14*	0.24**	0.36**	0.07 (0.179)	0.58**	0.38**
Social relationships	0.04 (0.322)	−0.04 (0.324)	0.37**	0.09 (0.112)	0.28**	0.19**
Environment	0.21**	0.16*	0.30**	0.27**	0.44**	0.39**
Inpatient sample (N = 149)						
Physical health	0.14*	0.35**	0.30**	0.04 (0.300)	0.46**	0.37**
Psychological health	0.14 (0.05)	0.21**	0.32**	0.02 (0.386)	0.54**	0.35**
Social relationships	0.05 (0.278)	−0.11 (0.099)	0.32**	0.085 (0.153)	0.23**	0.16**
Environment	0.19**	0.11 (0.095)	0.26**	0.24**	0.38**	0.34**
Outpatient sample (N = 22)						
Physical health	0.42*	0.47*	0.60**	0.11 (0.311)	0.45*	0.56**
Psychological health	0.24 (0.140)	0.54**	0.59**	0.37*	0.77**	0.67**
Social relationships	0.01 (0.480)	0.43*	0.68**	0.13 (0.282)	0.56**	0.47*
Environment	0.22 (0.164)	0.52**	0.60**	0.52**	0.77**	0.67**

p values are displayed within parentheses for coefficients with $p > 0.05$. *$p < 0.05$; **$p < 0.01$.

the inpatient sample, we found correlation coefficients ranging from −0.11 to 0.54, showing a similar pattern to the total sample. In the outpatient sample, we found stronger correlations between the SURE-NL scores, particularly the SURE-NL "self-care" subscale score, and the WHOQoL-BREF domain scores, compared to the inpatient and total sample.

Discriminant Validity

The DASS-21 subscales had high reliabilities in our sample: "depression" Cronbach's α = 0.90; "anxiety" Cronbach's α = 0.83; and "stress" Cronbach's α = 0.89. Table 5 presents the Spearman's correlations between the SURE-NL total and subscale scores and the DASS-21 subscale scores. Correlation coefficients ranged from −0.18 to 0.01. Apart from the correlation between the SURE-NL "outlook on life" subscale and the DASS-21 "anxiety" subscale, all correlations were small and not statistically significant, indicating good discriminant validity.

SURE-NL Scores

The median total score for the 21-item SURE-NL for the total sample was 54 (IQR 48–58). Table 6 shows the total and subscale scores for the SURE-NL for the participants recruited in inpatient (N = 149) and outpatient

Table 5. Spearman's rho correlation coefficients for the SURE-NL total and subscale scores and DASS-21 subscale scores

	SURE-NL					
	substance use	self-care	relationships	material resources	outlook on life	total score
DASS-21						
Depression	−0.04 (0.595)	0.01 (0.882)	−0.05 (0.520)	−0.06 (0.462)	−0.11 (0.139)	−0.07 (0.347)
Anxiety	−0.08 (0.322)	−0.02 (0.832)	−0.04 (0.603)	−0.06 (0.430)	−0.18*	−0.10 (0.198)
Stress	−0.05 (0.490)	−0.04 (0.596)	−0.09 (0.255)	−0.06 (0.430)	−0.09 (0.270)	−0.09 (0.250)

p values are displayed within parentheses for coefficients with $p > 0.05$. *$p < 0.05$.

Table 6. SURE-NL scores comparisons between participants in inpatient and outpatient treatment

	Inpatient ($N = 149$), median (IQR)	Outpatient ($N = 22$), median (IQR)	Comparison	Effect size
Substance use	16 (14–18)	13.5 (12–16)	**u = 2211.50, *z* = 2.69, *p* < 0.01**	$r = 0.21$
Self-care	13 (11–15)	11 (7.75–13)	**u = 2215.50, *z* = 2.70, *p* < 0.01**	$r = 0.21$
Relationships	11 (11–12)	11 (9.5–12)	u = 1,891.50, $z = 1.25$, $p = 0.211$	$r = 0.10$
Material resources	8 (7–9)	8.5 (6–9)	u = 1,638.50, $z = -0.002$, $p = 0.998$	$r = 0.00$
Outlook on life	7 (6–8)	6 (4–8.25)	u = 1,997.50, $z = 3.32$, $p = 0.092$	$r = 0.25$
Total score	54 (50–58)	47 (41.5–56.25)	**u = 2171.50, *z* = 1.69, *p* < 0.05**	$r = 0.13$

Significant values ($p < 0.05$) appear in bold.

treatment ($N = 22$). Those in inpatient treatment scored significantly higher than those in outpatient treatment on the SURE-NL subscales "self-care" and "substance use" and – consequently – on the SURE-NL total score.

Discussion

In this study, we examined the validity and reliability of the Dutch translation of the SURE, the SURE-NL, for measuring indicators of personal recovery and recovery capital in people with SUD in a Dutch-speaking population of patients in inpatient and outpatient SUD treatment in Belgium. The SURE was originally developed in English by Neale et al. [11] as a 21-item instrument with a total score and five subscales, based on factor analysis with Promax oblique rotation. The German translation by Reichl et al. [18], using a factor analysis with Oblimin oblique rotation, had 20 items and three factors. We performed a CFA, showing that the original 5-factor structure as proposed by Neale et al. [11] had an acceptable fit for our data.

Observed correlations between the SURE-NL subscales ranged from 0.19 to 0.54. The "relationships" and the "outlook on life" subscales were strongly inter-correlated ($r = 0.54$), suggesting considerable overlap between these two concepts. A possible explanation for this is that having a supportive network has a positive impact on quality of life, which is assessed in the "outlook on life" subscale. In comparison, inter-scale correlations in the original English SURE ranged from 0.40 to 0.70, and those in the German version from 0.39 to 0.54 [11, 18]. All subscales have a strong positive correlation with the total score (correlations between 0.51 and 0.77), indicating that the SURE-NL subscales assess a common underlying concept. Therefore, a 1-factor structure was considered, but the CFA indicated a better fit for the 5-factor model. We used both subscale scores, which provide detailed information on specific dimensions of recovery, and the total score, which offers a measure of the overall construct of recovery.

Internal consistencies of the SURE-NL subscales "material resources", "substance use," and "relationships" were relatively low (Cronbach's $\alpha < 0.70$), while the

SURE-NL "self-care" and "outlook on life" subscales had good internal consistency (Cronbach's α = 0.76). Internal consistency of the total score of the SURE-NL was good: Cronbach's α = 0.83. For the original SURE, Cronbach's α of the total score was 0.92, and for the five subscales, this ranged from Cronbach's α = 0.68 ("material resources") to Cronbach's α = 0.87 ("outlook on life") [11]. The three subscales of the German translation by Reichl et al. (2023) all showed good internal consistency: "psychological and physical well-being" (8 items, Cronbach's α = 0.86), "daily functioning" (7 items, Cronbach's α = 0.76), and "substance use" (5 items, Cronbach's α = 0.85), and had a Cronbach's α of 0.89 for the total score [18]. These higher internal consistencies found in the German version might be (partly) explained by the larger number of items of each subscale.

Regarding concurrent validity, we found mostly positive correlations of the SURE-NL total and subscale scores with the WHOQoL-BREF subscales. However, for the total and inpatient sample, these correlations were smaller than those found by Neale et al. [11], especially for the SURE-NL "substance use" and "material resources" subscales, correlation coefficients were small. This is likely due to the treatment setting, which strongly influences access to substances, housing, and money, which are topics assessed in these subscales. This is further supported by the stronger correlations between the SURE-NL and the WHOQoL-BREF subscales in our outpatient sample, which were similar to those found in the outpatient study by Neale et al. [11]. Additionally, our findings show that the SURE-NL is not correlated with the DASS-21, a measure for psychopathology, suggesting that these constructs are distinct from recovery, supporting the discriminant validity of the SURE-NL. The only small, significant, negative correlation we found was between the SURE-NL "outlook on life" subscale and the DASS-21 "anxiety" subscale, which might be due to symptoms of anxiety influencing people's response to questions such as "I have felt positive."

In our entire sample, the median total score on the 21-item SURE-NL was 54 (IQR 48–58). Neale et al. [11] found a median total score of 41.8 (min. 22 – max. 63) in their sample [11]. Although it was suggested by Neale et al. [11] that questions such as "having stable housing" and "managing money well" would be influenced by the structure of an inpatient setting, we found no significant difference in the magnitude of the scores between participants in inpatient and outpatient treatment for the SURE-NL "material resources" subscale. Furthermore, we observed that scores on the "substance use" subscale of the SURE-NL were significantly higher for participants in inpatient treatment than for those in outpatient treatment, indicating less substance use than those in inpatient treatment. This is to be expected considering that abstinence is required in inpatient settings and motivation for change is a prerequisite to start treatment. We also found a significant difference in SURE-NL "self-care" scores between participants in inpatient treatment and those in outpatient treatment, which is likely due to inpatient participants scoring higher on questions such as "eating a good diet" and "having a good daily routine." Finally, the higher scores for participants in inpatient treatment on the SURE-NL "substance use" and "self-care" subscales seem to explain the significantly higher total scores on the SURE-NL for participants in inpatient treatment compared to those in outpatient treatment, even though both groups were in the same early stage of treatment.

The relatively low number of outpatients compared to inpatient participants and the heterogeneity of the participant sample included in this study is a limitation to consider. Although the SURE was initially developed for use in community-based services, most of our participants (87.2%) were in inpatient treatment. This is important since inpatients were not included in the original study because the SURE was developed for an outpatient population and several items of the SURE assess aspects that are likely influenced by treatment settings, for example, diet, daily routine, and housing. Therefore, it is important to investigate the impact of the treatment setting on the SURE-NL scores carefully, and to interpret results in their context. The CFA was conducted on a sample that included both inpatient and outpatient participants. While it would be beneficial to perform a multigroup analysis to establish subgroup measurement invariance, this would require a larger sample, particularly for the outpatient subgroup. The relatively small sample (N = 171) is another limitation to consider. Validity testing guidelines vary, suggesting respondent-to-item ratios anywhere between 5:1 and 30:1, with our sample size being toward the lower end of this spectrum with a respondent-to-item ratio of 8:1 [21]. Although our choice for classical test theories over IRT analyses limits the depth of item-level analysis, our use of CFA, reliability, and validity testing still provided a solid assessment of the psychometric properties of the SURE-NL, which can be compared to previous research. On the other hand, the investigated sample was diverse in terms of substances used, age, and education level. To the best of our knowledge, this study is the first to offer a psychometrically validated instrument for evaluating the recovery process in people with SUD in Dutch, which was developed with extensive service user input. It is only the second psychometrically validated translation of the SURE

available, and the first to investigate its use in a sample of participants that were predominantly in inpatient treatment.

Overall, the original 5-factor structure of the SURE showed an acceptable fit for our data. While our findings showed acceptable to good reliability and validity for the SURE-NL "self-care," "relationships," and "outlook on life" subscales and the total score, this was not the case for the "substance use" and "material resources" subscales. This is probably related to the fact that our sample predominantly consisted of participants in inpatient treatment settings, while the SURE was developed for use in an outpatient population. Although treatment setting seems to have an impact on the results, especially on some of the subscale scores, based on our findings, we conclude that the SURE-NL is appropriate for use in Dutch-speaking treatment-seeking people with SUD, including inpatients, but subscales should be used and interpreted with caution. Future research should assess the SURE-NL in a larger sample of Dutch and Belgian outpatients, and the development of a separate version for use in inpatient settings, for which the results of the current study can provide relevant input, should be considered.

Statement of Ethics

This study protocol was reviewed and approved by the UZ Brussel Ethics Committee on 11 May 2022, approval No. 1432022000071. Written informed consent was obtained from all participants.

Conflict of Interest Statement

Wim van den Brink reports a consulting/advisory relationship with Takeda Pharmaceutical Company Limited, Camurus AB, and Clearmind. All other authors have no conflicts of interest to declare.

Funding Sources

This work was supported by a grant from the Belgian Science Policy Office (BELSPO), research project DR/90. BELSPO had no role in the design, data collection, data analysis, and reporting of this study.

Author Contributions

C.M.: conceptualization, formal analysis, investigation, and writing – original draft; W.v.d.B.: conceptualization and writing – original draft; A.Z.: investigation and writing – review and editing; F.I.A.M. and C.D.R.: writing – review and editing; D.D.: formal analysis and writing – review and editing; W.V.: conceptualization, writing – review and editing, and supervision; C.L.C.: conceptualization, writing – original draft, and supervision. All authors approved the final manuscript.

Data Availability Statement

The data that support the findings of this study are not publicly available due to their containing information that could compromise the privacy of the research participants but are available from the first author, C.M., upon reasonable request.

References

1 Degenhardt L, Hall W. Extent of illicit drug use and dependence, and their contribution to the global burden of disease. Lancet. 2012;379(9810):55–70. https://doi.org/10.1016/S0140-6736(11)61138-0

2 Castelpietra G, Knudsen AKS, Agardh EE, Armocida B, Beghi M, Iburg KM, et al. The burden of mental disorders, substance use disorders and self-harm among young people in Europe, 1990-2019: findings from the Global Burden of Disease Study 2019. Lancet Reg Health Eur. 2022;16:100341. https://doi.org/10.1016/j.lanepe.2022.100341

3 Tiffany ST, Friedman L, Greenfield SF, Hasin DS, Jackson R. Beyond drug use: a systematic consideration of other outcomes in evaluations of treatments for substance use disorders. Addiction. 2012;107(4):709–18. https://doi.org/10.1111/j.1360-0443.2011.03581.x

4 Alves P, Sales C, Ashworth M. Does outcome measurement of treatment for substance use disorder reflect the personal concerns of patients? A scoping review of measures recommended in Europe. Drug Alcohol Depend. 2017;179:299–308. https://doi.org/10.1016/j.drugalcdep.2017.05.049

5 De Maeyer J, Vanderplasschen W, Broekaert E. Exploratory study on drug users' perspectives on quality of life: more than health-related quality of life? Soc Indic Res. 2009;90(1):107–26. https://doi.org/10.1007/s11205-008-9315-7

6 Kiluk BD, Fitzmaurice GM, Strain EC, Weiss RD. What defines a clinically meaningful outcome in the treatment of substance use disorders: reductions in direct consequences of drug use or improvement in overall functioning? Addiction. 2019;114(1):9–15. https://doi.org/10.1111/add.14289

7 Trujols J, Portella MJ, Iraurgi I, Campins MJ, Siñol N, de Los Cobos JP. Patient-reported outcome measures: are they patient-generated, patient-centred or patient-valued? J Ment Health. 2013;22(6):555–62. https://doi.org/10.3109/09638237.2012.734653

8 Churruca K, Pomare C, Ellis LA, Long JC, Henderson SB, Murphy LED, et al. Patient-reported outcome measures (PROMs): a review of generic and condition-specific measures and a discussion of trends and issues. Health Expect. 2021;24(4):1015–24. https://doi.org/10.1111/hex.13254

9 Migchels C, Zerrouk A, Crunelle CL, Matthys F, Gremeaux L, Fernandez K, et al. Patient Reported Outcome and Experience Measures (PROMs and PREMs) in substance use disorder treatment services: a scoping review. Drug Alcohol Depend. 2023;253:111017. https://doi.org/10.1016/j.drugalcdep.2023.111017

10 Neale J, Strang J. Philosophical ruminations on measurement: methodological orientations of patient reported outcome measures (PROMS). J Ment Health. 2015;24(3):123–5. https://doi.org/10.3109/09638237.2015.1036978

11 Neale J, Vitoratou S, Finch E, Lennon P, Mitcheson L, Panebianco D, et al. Development and validation of 'SURE': a Patient Reported Outcome Measure (PROM) for recovery from drug and alcohol dependence. Drug Alcohol Depend. 2016;165:159–67. https://doi.org/10.1016/j.drugalcdep.2016.06.006
12 Brophy H, Dyson M, Katherine RK. Concept analysis of recovery from substance use. Int J Ment Health Nurs. 2023;32(1):117–27. https://doi.org/10.1111/inm.13066
13 Bellaert L, Van Steenberghe T, De Maeyer J, Vander Laenen F, Vanderplasschen W. Turning points toward drug addiction recovery: contextualizing underlying dynamics of change. Addiction Res Theor. 2022;30(4):294–303. https://doi.org/10.1080/16066359.2022.2026934
14 Laudet AB. What does recovery mean to you? Lessons from the recovery experience for research and practice. J Subst Abuse Treat. 2007;33(3):243–56. https://doi.org/10.1016/j.jsat.2007.04.014
15 Leamy M, Bird V, Le Boutillier C, Williams J, Slade M. Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. Br J Psychiatry. 2011;199(6):445–52. https://doi.org/10.1192/bjp.bp.110.083733
16 Best D, Sondhi A, Brown L, Nisic M, Nagelhout GE, Martinelli T, et al. The strengths and barriers recovery scale (SABRS): relationships matter in building strengths and overcoming barriers. Front Psychol. 2021;12:663447. https://doi.org/10.3389/fpsyg.2021.663447
17 Vanderplasschen W, Martinelli T, Bellaert L, Nagelhout G, Mheen D. Herstel na een verslaving aan illegale drugs: resultaten van de Life in Recovery survey in Nederland en Vlaanderen. Verslaving en herstel. 2021;1(1):9–19.
18 Reichl D, Enewoldsen NM, Berking M, Fuhrmann LM, Lang C, Saur S, et al. Psychometrische Evaluation der deutschen version des substance use recovery evaluator (SURE). Diagnostica. 2023;69(2):51–61. https://doi.org/10.1026/0012-1924/a000301
19 Skevington SM, Lotfy M, O'Connell KA; WHOQOL Group. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res. 2004;13(2):299–310. https://doi.org/10.1023/B:QURE.0000018486.91360.00
20 Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the depression anxiety stress scales (DASS) with the beck depression and anxiety inventories. Behav Res Ther. 1995;33(3):335–43. https://doi.org/10.1016/0005-7967(94)00075-u
21 Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi J Anaesth. 2017;11(Suppl 1):S80–9. https://doi.org/10.4103/sja.SJA_203_17
22 Beaufort IN, De Weert-Van Oene GH, Buwalda VAJ, de Leeuw JRJ, Goudriaan AE. The depression, anxiety and stress scale (DASS-21) as a screener for depression in substance use disorder inpatients: a pilot study. Eur Addict Res. 2017;23(5):260–8. https://doi.org/10.1159/000485182
23 Vanderplasschen W, Van Baelen L, Matthys F, Crunelle CL. Outcome Measurement and Evaluation as a Routine practice in alcohol and other drug services in Belgium (OMER-BE). 2020. [cited 2023 December 11] Available from: https://www.belspo.be/belspo///fedra/proj.asp?l=de&COD=DR%2F90#docum
24 Vanderplasschen W, Vandevelde S, Broekaert E. Therapeutic communities for treating addictions in Europe: evidence, current practices and future challenges. Luxembourg: Publications Office of the European Union; 2014. https://doi.org/10.2810/25291
25 LG. LimeSurvey: an open source survey tool. (n.d.).
26 Team RCR. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2024.
27 Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model: A Multidiscip J. 1999;6(1):1–55. https://doi.org/10.1080/10705519909540118
28 Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012