








ORIGINAL ARTICLE

# Measuring personal recovery in people with a psychotic disorder based on CHIME: A comparison of three validated measures

Jelle Sjoerd Vogel,<sup>1,2,3</sup>  Jojanneke Bruins,<sup>1,3</sup>  Levi Halbersma,<sup>1</sup> Rianne Janine Lieben,<sup>1</sup> Steven de Jong,<sup>1,3</sup>  Mark van der Gaag,<sup>4,5,6</sup>  and Stynke Castelein<sup>1,2,3</sup> 

<sup>1</sup>Lentis Psychiatric Institute, <sup>2</sup>Faculty of Behavioral and Social Sciences, University of Groningen, <sup>3</sup>Rob Giel Research Center, University Medical Center Groningen, University of Groningen, Groningen, <sup>4</sup>Department of Clinical Psychology, VU University, <sup>5</sup>Amsterdam Public Health Research Institute, Amsterdam, and <sup>6</sup>Department of Psychosis Research, Parnassia Psychiatric Institute, The Hague, The Netherlands

**ABSTRACT:** *Living well in spite of residual symptoms of mental illness is measured with the construct of personal recovery. The CHIME framework might be suitable to evaluate personal recovery measures and guide instrument choice. Three validated measures were evaluated in Dutch patients with a psychotic disorder (N = 52). We compared the Recovery Assessment Scale (RAS), the Mental Health Recovery Measure (MHRM), and the Netherlands Empowerment List (NEL). The measures were assessed on six criteria: content validity (based on CHIME), convergent validity with a social support measure, internal consistency, floor and ceiling effects, item interpretability, and ease of administration. The MHRM scored high on content validity with a balanced distribution of items covering the CHIME framework. The MHRM and the NEL showed moderate convergent validity with social support. In all three measures, internal consistency was moderate and floor and ceiling effects were absent. The NEL scores demonstrated a high degree of item interpretability. Ease of administration was moderate for all three measures. Finally, the CHIME framework demonstrated good utility as a framework in guiding instrument choice and evaluation of personal recovery measures. The MHRM showed the best overall result. However, differences between measures were minimal. Generalization of the results is limited by cultural and linguistic factors in the assessment for the subjective measures (i.e. content validity and item interpretability). The broad and multidimensional construct of personal recovery might lead to ambiguous interpretations. Scientific consensus on a well-defined personal recovery construct is needed.*

**KEY WORDS:** *mental health recovery, outcome assessment, psychometrics, psychosis, Schizophrenia spectrum and other psychotic disorders.*

**Correspondence:** J.S. Vogel, Hereweg 80, 9725 AG Groningen, The Netherlands. Email: js.vogel@lentis.nl

**Declaration of conflict of interest:** There are no conflicts of interest to disclose.

**Authorship statement:** All authors listed meet the authorship criteria according to the latest guidelines of the International Committee of Medical Journal Editors, and all authors agree with the manuscript.

Jelle Sjoerd Vogel, MSc.

Jojanneke Bruins, PhD.

Levi Halbersma, MSc.

Rianne Janine Lieben, BSc.

Steven de Jong, PhD.

Mark van der Gaag, PhD.

Stynke Castelein, PhD.

Accepted February 11 2020.

© 2020 The Authors. *International Journal of Mental Health Nursing* published by John Wiley & Sons Australia, Ltd on behalf of Australian College of Mental Health Nurses Inc.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

## INTRODUCTION

In recent years, the patient movement raised awareness on the importance of personal recovery (Liberman & Kopelowicz 2005; Roe *et al.* 2011). Personal recovery is defined by Anthony as a ‘deeply personal, unique process of changing one’s attitudes, values, feelings, goals, skills and/or roles. It is a way of living a satisfying, hopeful, and contributing life even with limitations caused by the illness. Recovery involves the development of new meaning and purpose in one’s life as one grows beyond the catastrophic effects of mental illness’ (Anthony 1993). Several studies advocate for the implementation of personal recovery in nursing care (Jacob *et al.* 2017; Noiseux & Ricard 2008). The growing amount of research on personal recovery-focused interventions emphasizes the increasing need for the evaluation of personal recovery measures.

Recently, the CHIME framework for recovery has received increased attention as a way to operationalize recovery (Leamy *et al.* 2011). This framework is based on a qualitative review of 115 studies. A narrative approach resulted in the five themes of the CHIME framework: *Connectedness* (support from others and being part of the community), *Hope* (a positive view on the future and motivation to change), *Identity* (building a positive identity and overcoming stigma), *Meaning* (developing meaningful roles and activities, quality of life, and spirituality), and *Empowerment* (gaining control over life and focus upon strengths; Leamy *et al.* 2011). The importance of CHIME is widely endorsed in the literature (van Weeghel *et al.* 2019) which makes the framework suitable for evaluating personal recovery measures.

Self-report instruments for personal recovery, with straightforward interpretation such as those used in depression (e.g. *Inventory of Depressive Symptomatology* (IDS)), could serve evaluation of clinical care as well as research. However, personal recovery is characterized by its subjective nature and unique experiences (Wood & Alsawy 2018). Therefore, personal recovery measures contain broadly interpretable items so that a measure can cover a large variety of personal recovery experiences for patients with divergent levels of (cognitive) functioning (Bellack 2006). These aspects make personal recovery measures prone to the bandwidth–fidelity problem. This problem refers to the trade-off between obtaining a score with high fidelity and narrow bandwidth on a specific trait versus obtaining a score that covers a broad bandwidth with less fidelity (Streiner & Norman 2008). Therefore, developing a generic measure

with a broad bandwidth comes at the expense of specificity: mean scores on studies that use a personal recovery measure as the primary outcome can generate multiple interpretations. As an alternative approach, the dimensions of the CHIME framework could serve as single constructs resulting in more specificity compared to the multidimensional construct of personal recovery.

Furthermore, subscales in measures are commonly identified with factor analyses. However, in heterogeneous constructs, factor analyses do not always produce subscales with sufficient content validity as they rely on internal consistency of subscales. In other words, while factor analysis retains items which correlate highly with one another and thus improves internal consistency, it may also cause items to be dropped which have a lower correlation but are important for content validity (Streiner & Norman 2008). Given how CHIME was developed based on a qualitative review of the literature, the dimensions of the CHIME framework may produce subscales based on content validity rather than factor analysis.

In this study, we evaluate the Recovery Assessment Scale (RAS) (Corrigan *et al.* 1999), the Mental Health Recovery Measure (MHRM) (Young & Bullock 2005), and the Netherlands Empowerment List (NEL) (Boevink *et al.* 2016). The choice of measures was based on different qualities of the measures. The RAS was included because it is one of the earliest developed measures and is currently the most widely used scale for measuring personal recovery (Salzer & Brusilovskiy 2014). However, the RAS was not validated for the Dutch-speaking population. Therefore, the MHRM was included as it was the first validated scale for measuring personal recovery in the Dutch mental health-care population (van Nieuwenhuizen *et al.* 2014). However, both the RAS and MHRM are developed in Anglo-Saxon countries and later translated in other languages. Translated measures may have limited generalizability to their targeted population and reliability measures, such as Cronbach’s alpha, or floor and ceiling parameters might perform different in their country of development (Streiner & Norman 2008). Validation of the reliability of measures in a specific language is therefore needed. We therefore included the NEL, as it was originally developed in Dutch and validated in the Netherlands.

Personal recovery is a highly subjective construct in which language and cultural elements likely play an important role (Brijnath 2015; Slade 2012). Consequently, we expect the NEL to perform better on measures of applicability as this measure will be evaluated in the same country as it was developed. Based on

prior research, we expect all three measures to perform adequately on construct validity, reliability, and applicability (Boevink *et al.* 2016; Burgess *et al.* 2011; Cavelti *et al.* 2012; van Nieuwenhuizen *et al.* 2014; Salzer & Brusilovskiy 2014; Shanks *et al.* 2013; Sklar *et al.* 2013).

## AIMS

In this study, we hypothesize that a recovery measure shows higher applicability scores when applied in the country of origin, compared to recovery measures from foreign countries. Therefore, we compare the RAS, the MHRM, and the NEL for measuring personal recovery in a Dutch population of patients with a psychotic disorder. Measures will be exploratory assessed on aspects of construct validity (content and convergent validity) and reliability (Cronbach's alpha and floor and ceiling effects) and explanatory on applicability for patients with psychosis (interpretability and ease of administration). Furthermore, we explore the CHIME as framework for analysing content validity.

## METHODS

### Sample

A convenience sample was recruited within Lentis Psychiatric Institute (Groningen, The Netherlands) from October 2015 to February 2017. Patients with a psychotic disorder (schizophrenia spectrum: 295.xx, 297.1, 298.80, 298.90) (DSM IV, American Psychiatric Association 2000), age  $\geq 18$ , were eligible for inclusion. In total, 53 patients were recruited for this study. One patient was excluded due to missing data. Therefore, 52 patients were included in the analysis. A description of the sample is presented in Table 1.

### Procedures and materials

All included personal recovery measures are self-report questionnaires and use a Likert scale (range: 1 = strongly disagree to 5 = strongly agree) with higher scores representing more personal recovery.

#### *Recovery Assessment Scale*

The RAS consists of 24 items (range 24-120). Corrigan *et al.* (1999) conducted a factor analysis resulting in five factors: *Personal confidence and hope*, *Willingness to ask for help*, *Goal and success orientation*, *Reliance on others*, and *No domination by symptoms* (Data S1).

Good psychometric properties were found in the original English measure, with Cronbach's alpha ranging between  $\alpha = 0.76$  and  $\alpha = 0.97$  (Salzer & Brusilovskiy 2014). Test-retest reliability ranged from  $r = 0.65$  to  $r = 0.88$ . Three out of six RCT's found the RAS to be sensitive to change in different psychiatric disorders (Salzer & Brusilovskiy 2014). In four reviews, the RAS received good evaluations for convergent validity and ease of administration (Burgess *et al.* 2011; Cavelti *et al.* 2012; Shanks *et al.* 2013; Sklar *et al.* 2013). For this study, the RAS was translated to Dutch using forward-backward translation with a native English speaker (Data S2).

#### *Mental Health Recovery Measure*

The MHRM was developed based on a recovery model by Young and Ensing (2005). It consists of 30 items with a scoring range between 30 and 150. In two

**TABLE 1** Description of the sample (N = 52)

Item	Mean (SD)
Age	42.3 (12.69)
Illness duration (years)	15.1 (11.05)
	N (%)
Female/Male	12 (23.1)/ 40 (77.9)
Primary diagnosis	
Schizophrenia	34 (65.4)
Psychotic disorder NOS	8 (15.4)
Schizoaffective disorder	7 (13.5)
Other	2 (3.8)
Unknown (missing data)	1 (1.9)
Type of care	
Functional Assertive Community Treatment	28 (53.8)
Functional Assertive Community	10 (19.2)
Treatment + sheltered housing	
Hospital admitted	4 (7.7)
Rehabilitation care	10 (19.2)
Self-reported psychosis	
1	10 (19.2)
2	10 (19.2)
3	8 (15.4)
>3	11 (21.2)
Unknown	13 (25)
Occupation	
Paid work	8 (15.1)
Voluntary work	21 (39.6)
Study	3 (5.7)
Other (e.g.: occupational therapy, domestic work)	21 (39.6)
Education <sup>†</sup>	
Primary education	6 (11.6)
Lower secondary education	14 (26.9)
Upper secondary education	24 (46.2)
Bachelor or Master	8 (15.3)

<sup>†</sup>International Standard Classification of Education.

reviews, the MHRM received good evaluations for internal consistency and ease of administration. Convergent validity was measured with constructs of empowerment, resilience, and community living. The correlations varied from 0.57 to 0.75 (Cavelti *et al.* 2012; Sklar *et al.* 2013). The Dutch version of the MHRM was used for this study. Exploratory factor analyses conducted on the Dutch MHRM reduced the original seven subscales to three (van Nieuwenhuizen *et al.* 2014): *Self-empowerment*, *Learning and new potentials*, and *Spirituality* (Data S1). Cronbach's alphas for the three factors in the Dutch MHRM ranged from  $\alpha = 0.86$  to  $\alpha = 0.94$  (van Nieuwenhuizen *et al.* 2014).

#### *Netherlands Empowerment List*

The NEL was developed in the Netherlands and based on the results of a conceptual study of empowerment. In this study, concept mapping was used with participation of 56 patients with severe mental illness (Boevink *et al.* 2017). The 40-item measure has a scoring range from 40 to 200. The NEL consists of six subscales: *Social support*, *Professional help*, *Connectedness*, *Confidence and purpose*, *Self-management*, and *Caring community* (Data S1). Good psychometric properties were found with a Cronbach alpha of  $\alpha = 0.94$  and a test-retest reliability of 0.79 (ICC). The NEL showed sensitivity to change in a two-year randomized controlled trial that evaluated a personal recovery intervention for people with severe mental illness (Boevink *et al.* 2016).

#### *Social Support List 12 – Interactions*

Social support is recognized as a construct closely related to personal recovery (Salzmann-Erikson 2013; Wood & Alsawy 2018). Social support is embedded in the CHIME framework through the *Connectedness* dimension. Moreover, support from peers, friends, or family benefits all aspects of the framework to a certain extent (as noted in subdimensions of CHIME such as *Hope inspiring relationships* or *Meaningful life and social roles*). The importance of social support is also reflected in studies of personal recovery-focused interventions, which aim to strengthen the social network (Castelein *et al.* 2015; Vogel *et al.* 2019). Therefore, correlation of a personal recovery measure with a social support measure could serve to strengthen theoretical support for the construct validity.

To assess convergent validity with social support, we used the Social Support List 12-Interactions (SSL12-I) (van Eijk *et al.* 1994). The SSL12-I was validated for

use in the geriatric population (Kempen & van Eijk 1995). The 12-item measure (range: 12–48) contains three subscales: *Everyday support*, *Esteem support*, and *Support in problem situations*. The SSL12-I showed satisfactory psychometric properties with all subscales demonstrating an internal reliability coefficient of  $r \geq 0.70$ . The mean interitem correlations were  $r = 0.29$  for the complete measure and  $r \geq 0.37$  for the subscales.

#### *Administration*

Each personal recovery measure was completed along with an item on 'ease of administration' using the same 5-point Likert scale as the questionnaires. Furthermore, participants provided information about demographic variables, such as highest level of completed education and diagnosis.

#### *Ethical considerations*

This study is completed in accordance with the declaration of Helsinki (World Medical Association (WMA) 2013), and all participants provided written informed consent. Participation was on voluntary basis with no compensation.

#### **Analyses**

Measures were analysed on six criteria: content validity, convergent validity, internal consistency, floor and ceiling effects, interpretability, and ease of administration. Each criterion was scored on a scale from 0 to 2 (0 = inadequate performance, 1 = moderate performance, and 2 = adequate performance). All criteria and their scoring are further specified below.

#### *Power analysis*

The number of participants needed for this study was based on the formula that is described in Streiner and Norman (2008). Based on earlier research, we expected to estimate values of Cronbach's alpha close to  $\alpha = 0.90$  in the current study (Boevink *et al.* 2016; van Nieuwenhuizen *et al.* 2014; Salzer & Brusilovskiy 2014). With  $\alpha = 0.90$  (95% CI = 0.85 and 0.95) and >25 items per scale, the power analysis resulted in 50 persons.

#### *Content validity*

We operationalized content validity by matching each item of each questionnaire to one of the five dimensions of the CHIME framework. The individual items were mapped to the CHIME framework by three criteria.

First, an item was considered valid when it reflected at least one of the five dimensions of the framework. Second, items that did not reflect any of the CHIME dimensions were considered superfluous. Content validity was based on the premise that all aspects of CHIME were considered equally important. A measure was therefore considered an adequate representation of personal recovery when all five dimensions of the CHIME framework were represented, and the items of the questionnaire were evenly distributed across all dimensions (Haynes *et al.* 1995). The mean number of valid items and their deviation from the mean (variance ( $\sigma$ )) across the different dimensions of CHIME was used as a measure for the total fit of the questionnaire to the framework. If the percentage of redundant items was  $\leq 10\%$  with a total variance of  $\sigma \leq 30$ , content validity was considered adequate (i.e. two points). Furthermore, content validity was considered moderate with  $>10\%$  and  $\leq 15\%$  redundant items and  $\sigma > 30$  and  $\sigma \leq 60$  and inadequate with  $>15\%$  redundant items and  $\sigma > 60$ . Content validity was assessed by three authors (JSV, JB, and RJL). The inter-rater agreement was analysed with ICC estimates (two-way mixed) and their 95% CI based on the mean and absolute agreement, resulting in a fair, good, or excellent agreement (Orwin 1994). A fourth author (SC) could be consulted in case discrepancies could not be solved in a consensus meeting. Having allocated each item to one of the CHIME domains resulted in an alternative subscale division for each instrument, with each of the CHIME domains serving as a subscale. Significant correlations between these new subscales and acceptable values of Cronbach's alpha ( $\alpha = 0.70$ ) would strengthen the validity of using CHIME as a basis for content validity analysis.

#### *Convergent validity*

Convergent validity for social support was operationalized by calculating a Pearson product-moment correlation coefficient between the recovery measures and the related construct 'social support', as measured with the SSL12-I. Correlations were considered small if  $r = 0.10$ , medium if  $r = 0.30$ , or high if  $r = 0.50$  (Cohen 1998). Convergent validity was considered adequate if  $r \geq 0.50$  with  $P < 0.05$ , moderate if  $r \geq 0.30$ , and  $< 0.50$  with  $P < 0.05$  or inadequate if  $P > 0.05$ . Additional analyses were performed on subscales that primarily focused on connectedness (i.e. *Reliance on others* (RAS) and *Social support* (NEL)). Furthermore, Pearson correlations between the three measures were calculated to evaluate whether the same construct was measured between the measures.

#### *Internal consistency*

Cronbach's alpha was used to analyse internal consistency on total measures and on original subscales as well as the alternative CHIME subscales that were created in this study. Values of  $\alpha < 0.70$  reflect poor internal consistency and values of  $\alpha > 0.90$  reflect redundancy of items (i.e. items have too much similarity) (Streiner & Norman 2008). Measures with  $\geq 14$  items produce acceptable alphas (i.e. more items result in a higher alpha; Cortina 1993). Therefore, we also examined the mean interitem correlation of the total measures. Total measures and subscales with an alpha below  $\alpha < .70$  were further evaluated on item level by calculating alpha if the respective item was deleted. A mean interitem correlation between  $r = 0.15$  and  $r = 0.20$  was considered adequate (Streiner 2003). Internal consistency was considered adequate if all of the following criteria were met:  $\alpha \geq 0.70$  and  $\leq 0.90$  with a mean interitem correlation between  $r = 0.15$  and  $r = 0.20$  on total and subscales. Internal consistency was considered moderate if one of these criteria were met and inadequate if none of these criteria were met.

#### *Floor and ceiling effects*

Floor and ceiling effects occur when  $>15\%$  of the respondents attain the lowest or the highest possible score, respectively (Terwee *et al.*, 2007). This is often the result in items which measure the extreme ends of the scales, that is which measure performance at the highest or lowest ends of the spectrum. This results in a negative impact on content validity, as participants at either extreme end cannot be distinguished from one another. The absence of floor and ceiling effects was considered adequate if  $\leq 15\%$  of the sample had either the highest or the lowest possible score, respectively (moderate  $>15\%$  and  $\leq 17.5\%$ , inadequate  $>17.5\%$ ).

#### *Item interpretability*

All items of the measures were screened on interpretability. Each item was independently assessed by three authors (JSV, JB, and RJL). Item interpretability was assessed on seven aspects: ambiguous wording; double-barrelled questioning; jargon; negative wording; lengthiness; complex sentence construction; abstract terms; and face validity (Streiner & Norman 2008). An item was considered adequate if no weak properties applied, moderate if only one weak property applied, and inadequate if  $\geq 2$  weak properties applied. Agreement on scores was then reached in a consensus meeting. This criterion was considered adequate if  $\leq 25\%$

of the items on a measure had weak properties, moderate if  $> 25\%$  and  $\leq 50\%$  of the items had weak properties, and inadequate if  $> 50\%$  of the items had weak properties.

#### *Ease of administration*

The ease of administration was measured with a single Likert scale item at the end of each measure ('this questionnaire was easy to complete') ranging from 1 = strongly disagree to 5 = strongly agree. The result is presented with descriptive statistics (means and SD's). A Friedman test was conducted to evaluate significant differences between the three measures. Ease of administration was considered adequate with a mean score of  $\geq 4$ , moderate with a mean score of  $\geq 3$  and  $< 4$ , and inadequate with a mean score of  $< 3$ . Patients with higher recovery scores might find it easier to complete the measure. To assess an association between ease of administration and personal recovery, a Pearson correlation was conducted on ease of administration with all three personal recovery measures.

All statistical analyses were performed using SPSS Statistics version 22 (IBM Corp 2013). In case of missing values, pairwise deletion was applied.

## RESULTS

### Content validity

A good agreement was found between the raters (ICC = 0.783, 95% CI: 0.667–0.857). Table 2 shows the outcomes of the content validity analysis. The MHRM showed the most optimal item distribution across CHIME ( $\sigma = 25.6$ ). The RAS scored high on *Empowerment* with 25% of the items on this dimension. Most items of the NEL were assigned to *Connectedness* (27.5%). *Hope and Optimism* was underrepresented in the NEL (7.5%).

With the content validity analysis, items of measures were mapped to the CHIME framework. This resulted in an alternative subscale division of the measures in which each domain of CHIME could serve as a subscale. Significant correlations were found between the measures mapped by their respective CHIME dimension. Only the RAS and MHRM did not correlate on *Connectedness*. Other correlations ranged from  $r(50) = 0.335$ ,  $P < 0.05$  to  $r(50) = 0.717$ ,  $P < 0.05$ . See Table S1 for the subscales based on CHIME that resulted from the content validity analysis. Cronbach's alpha was applied to measure the internal consistency of the alternative

CHIME subscales and ranged from  $\alpha = 0.442$  to  $\alpha = 0.836$ . Acceptable Cronbach's alphas ( $> 0.70$ ) were found for *Connectedness* on all three measures, for *Identity* on the MHRM and the NEL and for *Hope and Empowerment* on the NEL. See Table S2 for the internal consistency of all total and subscale measures.

#### *Convergent validity*

A significant medium correlation was found between the SSL12-I and both the MHRM ( $r(50) = 0.414$ ,  $P < 0.01$ ) and NEL ( $r(50) = 0.418$ ,  $P < 0.01$ ). The correlation between the SSL12-I and the RAS did not reach significance. Two subscales of the RAS and the NEL with a focus on connectedness were relevant for individual analyses. These subscales significantly correlated with the SSL12-I with, respectively, a medium correlation of  $r(50) = 0.406$ ,  $P < 0.01$  (*Reliance on others* (RAS)) and a high correlation of  $r(50) = 0.607$ ,  $P < 0.01$  (*Social support* (NEL)). All recovery measures correlated significantly with each other (RAS and MHRM  $r(50) = 0.697$ ,  $P < 0.01$ , RAS and NEL  $r(50) = 0.692$ ,  $P < 0.01$ , MHRM and NEL  $r(50) = 0.821$ ,  $P < 0.01$ ).

#### *Internal consistency*

All three measures showed acceptable homogeneity ( $\alpha \geq 0.70$ ) on total scores. The total score on the RAS ( $\alpha = 0.838$ ), as well as four out of five subscales of the RAS, showed an adequate internal consistency. The mean interitem correlation of the RAS was sufficient with  $r = 0.192$ . The subscales *No domination by symptoms* did not reach the lower boundary ( $\alpha = 0.671$ ). An analysis on alpha if item deleted showed that by deleting the item *My symptoms seem*

**TABLE 2** Content validity rating of the RAS, MHRM, and NEL items as measured with the CHIME framework

CHIME categories	RAS (24 items) items (%)	MHRM (30 items) items (%)	NEL (40 items) items (%)
Connectedness	4 (16.7)	4 (13.3)	11 (27.5)
Hope and optimism	3 (12.5)	4 (13.3)	3 (7.5)
Identity	3 (12.5)	7 (23.3)	4 (10.0)
Meaning and purpose	5 (20.8)	7 (23.3)	7 (17.5)
Empowerment	6 (25.0)	5 (17.7)	8 (20.0)
Item does not map	6 (12.5)	3 (10.0)	6 (17.5)
Variance	29.5	25.6	64.4

MHRM, Mental Health Recovery Measure; NEL, Netherlands Empowerment List; RAS, Recovery Assessment Scale.

to be a problem for shorter periods of time each time they occur alpha reached an adequate score ( $\alpha = 0.771$ ). The MHRM marginally exceeded the threshold for adequate internal consistency ( $\alpha = 0.903$ ) as well as for the mean interitem correlation of  $r = 0.245$ . The subscales *Self-empowerment* and *Learning and new potentials* showed an adequate alpha. The subscales *Spirituality* consist of only two items and showed an alpha of  $\alpha = 0.921$ . The NEL total score exceeded the upper threshold for both Cronbach's alpha ( $\alpha = 0.920$ ) and the mean interitem correlation ( $r = 0.223$ ). Five out of six subscales of the NEL have an adequate internal consistency. The subscale *Self-management* scored just under the threshold ( $\alpha = 0.628$ ). An analysis on alpha if item deleted did not result in  $\alpha > 0.70$  on a single-item deletion. See Table S2 for the internal consistency of all measures.

#### Floor and ceiling effects

None of the measures exceeded the 15% threshold for floor and ceiling effects. Floor effects were absent in total mean scores and subscale mean scores in all three measures. Also, ceiling effects were absent in total mean scores, but not in subscales. Ceiling effects on subscales ranged between 0% and 7.7% for the RAS, between 1.9% and 11.5% for the MHRM, and between 0% and 9.6% for the NEL.

#### Item interpretability

A little more than half (57.5%) of the items of the NEL was scored as straightforward to interpret by participants. Both the RAS and the MHRM scored low on item interpretability scores with >60% of items containing weak properties. For example, 'I feel at peace with myself (MHRM) was considered an ambiguous item. Table 3 shows the number and percentage of properties with respectively a good, fair, or poor rating.

#### Ease of administration

All three measures received a positive evaluation on ease of administration by the participants, with similar scores for the RAS ( $m = 3.71 \pm 0.87$ ), the MHRM ( $m = 3.51 \pm 1.05$ ), and the NEL ( $m = 3.67 \pm 0.96$ ). There were no significant differences ( $\chi^2 = 2.346$ ,  $P = 0.309$ ) between the measures. The RAS showed a significant positive correlation between the total recovery score and the ease of administration ( $r = 0.339$ ,  $P = 0.014$ ). Table 4 shows the summary of the six aspects that were measured.

## DISCUSSION

### Main findings

Of the three included personal recovery measures (RAS, MHRM, and NEL), only the MHRM showed good content validity with a balanced distribution of the items covering the complete CHIME framework. Unlike the NEL and MHRM, the RAS did not show convergent validity with the SSL12-I, despite the content validity analysis showing a reasonable number of items allocated to *Connectedness*. However, on subscale level, a medium significant correlation between the SSL12-I and the *Connectedness* subscale of the RAS (*Reliance on others*) was found. On interpretability, only the NEL achieved adequate scores on more than half of the items (57.5%). The MHRM showed the best result across all six criteria included in this evaluation due to the higher rating on content validity. This minimal difference does not suggest that the MHRM is convincingly better than the other measures. As noted in the introduction, the evaluation of measures resulted in an overall adequate score. Each instrument demonstrates a similar number of strengths and weaknesses, and as such, specific application may be the most useful guide in determining which instrument to use. If a greater focus on *Connectedness* is desired, the NEL offers the strongest properties. If *Empowerment* is of more interest, then we suggest using the RAS.

In our introduction, we hypothesized better performance on the NEL on measures of applicability. Although the NEL scored better on interpretability, it did not perform better on ease of administration compared to the other measures, thus rejecting our hypothesis.

**TABLE 3** Item interpretability scores of RAS, MHRM, and NEL

	RAS (24 items) Items (%)	MHRM (30 items) Items (%)	NEL (40 items) Items (%)
Good	9 (37.5)	7 (23.3)	23 (57.5)
Fair	11 (45.8)	16 (53.3)	11 (27.5)
Poor	4 (16.7)	7 (23.3)	6 (15.0)

Good: no weak properties; Fair: one weak property; Poor:  $\geq 2$  weak properties; MHRM, Mental Health Recovery Measure; NEL, Netherlands Empowerment List; RAS, Recovery Assessment Scale.

**TABLE 4** Summary of the evaluation of three personal recovery measures

	RAS	MHRM	NEL
Content validity	1	2	0
Convergent validity	0	1	1
Internal consistency	1	1	1
Floor and ceiling effects	2	2	2
Item interpretability	0	0	1
Ease of administration	1	1	1
Overall scoring	5	7	6

0, inadequate, 1, moderate, 2, adequate; MHRM, Mental Health Recovery Measure; NEL, Netherlands Empowerment List; RAS, Recovery Assessment Scale.

### Interpretation of the results

The CHIME framework was used in this study for analysing content validity. Items were allocated to one of the CHIME dimensions. For each measure, this resulted in alternative subscales based on the CHIME framework. Significant correlations were found between subscales of the same CHIME dimension. This outcome suggests that CHIME could be used for evaluation of personal recovery measures. Additionally, CHIME could also be used for the development of new scales based on its singular dimensions.

Similar to previous studies, we found evidence for internal consistency in all three measures (Boevink *et al.* 2017; van Nieuwenhuizen *et al.* 2014; Salzer & Brusilovskiy 2014). However, the MHRM and the NEL showed a Cronbach's alpha of >0.90, indicating redundant items in the measure. Our finding of a Cronbach's alpha of >0.90 on the subscale spirituality of the MHRM is in line with the study of van Nieuwenhuizen *et al.* (2014). Also, in line with our study, the study of Boevink *et al.* (2017) found an alpha of >0.90 on the total measure on the NEL. In the current study, acceptable alphas were found for all but one subscale of the NEL (*Self-management*,  $\alpha = 0.628$ ). This difference might be due to differences in sample characteristics: unlike the current study, the study of Boevink *et al.* (2017) was not limited to patients with psychosis. Furthermore, in the current study we found a possibility for a slight improvement on alpha in the RAS by deleting the item *My symptoms seem to be a problem for shorter periods of time each time they occur*. This will increase the internal consistency from  $\alpha = 0.671$  to  $\alpha = 0.771$  in the subscale *No domination by symptoms*. This finding is consistent with the theory on personal recovery, placing less emphasis on clinical symptoms.

The criteria for item interpretability were only partly met by the NEL and not by the RAS and the MHRM. Different populations might have a different understanding of the concept of personal recovery (Gopal & Henderson 2015) or the items or wording of a measure (Streiner & Norman 2008). For example, aspects of spirituality are differently conceptualized in diverse cultures (Brijnath 2015). As a possible result, interpretation of the items of the NEL was more straightforward for a Dutch team of researchers. This shows the importance of language in personal recovery measures. Furthermore, it reaffirms the influence of culture in the personal recovery construct (Brijnath 2015).

All measures showed moderate scores on ease of administration. In line with previous research, no differences were found in ease of administration (Cavelti *et al.* 2012; Sklar *et al.* 2013). Only the RAS showed a significant correlation with personal recovery scores, indicating that people with higher recovery scores find it easier to complete this questionnaire.

### Measuring personal recovery

The complexity and manifold dimensions of the construct of personal recovery are shown in this study and are discussed in several previous studies (Mathew *et al.* 2018; van Weeghel *et al.* 2019). For example, Liberman (2012) comments on the personal recovery construct by comparing it with self-efficacy measures that often leave 'much ambiguity about what is being measured' (Bandura 2006). Furthermore, Liberman (2012) argues that personal recovery is not a scientific construct and should therefore not be measured as an outcome in research. In an opinion paper, Bellack (2006) observed that the validity of self-assessment of personal experiences can be argued in patients with psychosis because of impaired reasoning and reality distortion (Bellack 2006; de Jong *et al.* 2018).

A large part of the complexity and ambiguity can be attributed to the wide variety of views on personal recovery. The lack of consensus on the personal recovery construct leads to many measures with different accents. Concrete examples are the subscale *Caring community* in the NEL, which is not seen in other personal recovery measures (Boevink *et al.* 2017). Furthermore, there is debate if personal recovery should be viewed as a process or an outcome (Law *et al.* 2012; Liberman 2012; Wood & Alsawy 2018). A recent scoping review of van Weeghel *et al.* (2019) supports the viewing of recovery as a process rather than an outcome. Also, Davidson (2012) argue's that supporting



patients who want to have a good life is at best a process with intermediate outcomes. The view of recovery as a process is reflected in measures that include stages of personal recovery. The Self-Identified Stages of Recovery (SISR; Andresen *et al.* 2003) and the Stages of Recovery Instrument (Andresen *et al.* 2006) are examples of questionnaires that measure stages in the process of recovery. Other questions are to what extent personal recovery is influenced by culture (Brijnath 2015; Ellison *et al.* 2018; Gopal & Henderson 2015) and which themes are most important in the personal recovery construct (e.g. social relationships and social activity, self-agency, or hope) (Ellison *et al.* 2018; Law *et al.* 2012). The CHIME framework offers a good starting point for consensus on these themes. However, even using the same framework can lead to different interpretations: in analysing content validity, based on CHIME, we came to different results than Shanks *et al.* (2013), who used the same framework. This review had much higher rates for the representation of the dimension *Hope and Optimism* (RAS: 17%, MHRM: 30%) and more items that did not match the CHIME framework in both the RAS (RAS: 29%, MHRM: 17%) compared to our evaluation.

### Limitations

Several limitations should be considered. First, in this review, content validity and item interpretability were only evaluated by researchers and a researcher/clinician. The absence of other stakeholders (e.g. patients, family) limits the validity of these analyses.

Second, the use of the CHIME framework for analysing content validity was confirmed by high correlations of items across measures mapped by the CHIME dimensions. However, a confirmatory factor analysis would show a more definite answer on the question whether CHIME is a suitable framework for analysing content validity. This was, however, not possible in this study due to a limited sample size. Furthermore, there are other frameworks available (Ellison *et al.* 2018) as well as adaptations on the CHIME framework (Bird *et al.* 2014; Stuart *et al.* 2017). This study only tested one framework for personal recovery and was therefore not able to analyse dimensions that are not part of CHIME such as *Difficulties* (Stuart *et al.* 2017). Although CHIME was shown to be a suitable framework for personal recovery in the current study, and several others, it is important to note that it was developed based on Western scientific literature (Leamy *et al.* 2011).

Third, all measures were administered to the participants in the same order. This might have biased the outcome on ease of administration as systematic differences could appear from a primacy or recency effect. A counterbalanced design, in which measures are administered in a random sequence, would therefore have been preferable.

Fourth, the evaluation of convergent validity was limited by only assessing it with social support. This was shown by measuring convergent validity in personal recovery on subscale level (i.e. subscales on *Connectedness*), which resulted in higher correlations compared to correlations on total measures. Due to the multidimensional nature of the personal recovery construct, evaluating convergent validity with other (CHIME-related) constructs such as empowerment, self-esteem, and hope, as well as divergent validity with self-stigma, loneliness, and depression, would provide additional insight.

### FUTURE RESEARCH

Personal recovery as defined by consumer views is still in need to transform vague, politically correct concepts into empirically reliable and valid scientific concepts (Lieberman 2012). Critical evaluation of the construct, especially on the dimensions to be included in the construct, is therefore needed. An important aspect in reaching this consensus is social validation (Lieberman & Kopelowicz 2005). With this form of validation, different stakeholders (patients, family, clinicians, researchers, and the broader public) are able to give input on the criteria for recovery. Furthermore, comprehensive construct validation could support the evidence for the personal recovery construct. However, it will not unambiguously prove the existence of such a construct (Streiner & Norman 2008).

The construct of personal recovery has gained much influence in mental health care. The call for the personal recovery model originated from a mental health-care system with a primary focus on symptom reduction (Anthony 1993). The consequences of a psychiatric disorder such as loneliness or loss of identity were underexposed in this medical paradigm. As a consequence, the patient movement advocated for a more holistic approach, considering the personal needs of patients with a psychiatric disorder. This raises the question whether the personal recovery construct should be used to guide mental health policy and practice rather than be used for evaluating mental health outcomes (Bellack 2006).

However, the emphasis placed on it by the patient movement demonstrates its great meaning to the clients themselves. As such, it is important that efforts are continued to further operationalize the construct. However, currently the idiosyncratic (Mathew *et al.* 2018) and culturally sensitive (Brijnath 2015) aspects of the construct impede the development of broad generic measures. In addition, models of recovery are primarily based on concepts of the Western world, which limits generalization to and international use of these measures in Non-Western cultures (Slade *et al.* 2012). Until more clarity is established on the construct of personal recovery, an option might be to disentangle the personal recovery construct to the CHIME dimensions or other common constructs (i.e. social support, hope, and goal orientation) and measure them as separate constructs (Law *et al.* 2012). Often these constructs have a longer history in scientific literature and have shown psychometric validity (Silverstein & Bellack 2008) as well as the before suggested social validity.

## CONCLUSIONS

A comparison of three personal recovery measures (RAS, MHRM, and NEL) on six criteria showed the MHRM to have the best results. However, there are several concerns with all personal recovery measures. The review process showed that measuring personal recovery is complex as there is little scientific agreement on the construct. In this regard, it should be considered that ‘not everything that can be counted counts, and not everything that counts can be counted’ (Cameron 1967). However, as an alternative to broad generic measures, unidimensional measures that are part of the personal recovery construct (e.g. hope, connectedness) could be used to measure specific aspects of personal recovery.

## RELEVANCE FOR CLINICAL PRACTICE

Personal recovery has gained an important role in mental health nursing care. In recent decades, many personal recovery measures were developed. This study evaluated three recovery measures (RAS, MHRM, and NEL). The MHRM showed the best result across all six criteria. However, the multidimensional and idiosyncratic nature of the personal recovery construct impedes a strong conclusion of a preferred measure. For use in mental health nursing practice, CHIME guided specific use of the NEL for a focus on *Connect- edness* and the RAS for a focus on *Empowerment*.

## ACKNOWLEDGEMENTS

The authors would like to thank Bas Bous and Anneke Zijlstra (Lentis Psychiatric Institute) for collecting the data for this research project and Courtney Bakker for her contribution as native English speaker in translating the Recovery Assessment Scale.

## REFERENCES

- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR*. Washington, DC: American Psychiatric Association.
- Andresen, R., Oades, L. G. & Caputi, P. (2003). The experience of recovery from schizophrenia: towards an empirically-validated stage model. *Australian and New Zealand Journal of Psychiatry*, 37, 586–594.
- Andresen, R., Caputi, P. & Oades, L. G. (2006). The stages of recovery instrument: development of a measure of recovery from serious mental illness. *Australian and New Zealand Journal of Psychiatry*, 40, 972–980.
- Anthony, W. A. (1993). Recovery from mental illness: The guiding vision of the mental health system in the 1990s. *Innovations and Research*, 2, 17–24.
- Bandura, A. (2006). *Self-Efficacy Beliefs of Adolescents*. New York: Information Age Publish.
- Bellack, A. S. (2006). Scientific and consumer models of recovery in schizophrenia: Concordance, contrasts, and implications. *Schizophrenia Bulletin*, 32, 432–442.
- Bird, V., Leamy, M., Tew, J., Le Boutillier, C., Williams, J. & Slade, M. (2014). Fit for purpose? validation of a conceptual framework for personal recovery with current mental health consumers. *The Australian and New Zealand Journal of Psychiatry*, 48, 644–653.
- Boevink, W., Kroon, H., van Vugt, M., Delespaul, P. & van Os, J. (2016). A user-developed, user run recovery program for people with severe mental illness: A randomised control trial. *Psychosis*, 8, 287–300.
- Boevink, W., Kroon, H., Delespaul, P. & Van Os, J. (2017). Empowerment according to persons with severe mental illness: Development of the Netherlands empowerment list and its psychometric properties. *Open Journal of Psychiatry*, 7, 18–30.
- Brijnath, B. (2015). Applying the CHIME recovery framework in two culturally diverse Australian communities: Qualitative results. *The International Journal of Social Psychiatry*, 61, 660–667.
- 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, 267–280.
- Cameron, W. B. (1967). *Informal Sociology, a Casual Introduction to Sociological Thinking*. New York: Random House.
- Castelein, S., Bruggeman, R., Davidson, L. & van der Gaag, M. (2015). Creating a supportive environment: Peer

- support groups for psychotic disorders. *Schizophrenia Bulletin*, 41, 1211–1213.
- Cavelti, M., Kvrjic, 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: The Journal of the Association of European Psychiatrists*, 27, 19–32.
- Cohen, J. W. (1998). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Corp, I. B. M. (2013). *IBM SPSS Statistics for Windows, version 22.0*. Armonk, NY: IBM corp.
- Corrigan, P. W., Giffort, D., Rashid, F., Leary, M. & Okeke, I. (1999). Recovery as a psychological construct. *Community Mental Health Journal*, 35, 231–239.
- Cortina, J. M. (1993). What is coefficient alpha? an examination of theory and applications. *Journal of Applied Psychology*, 78, 98–104.
- Davidson, L. (2012). Considering recovery as a process: Or, life is not an outcome. In: A. Rudnick (Ed). *Recovery of People with Mental Illness: Philosophical and Related Perspectives* (pp. 252–263). New York, NY: Oxford University Press.
- van Eijk, L. M., Kempen, G. I. & van Sonderen, F. L. (1994). A short scale for measuring social support in the elderly: The SSL12-I. *Tijdschrift voor Gerontologie en Geriatrie*, 25, 192–196.
- Ellison, M. L., Belanger, L. K., Niles, B. L., Evans, L. C. & Bauer, M. S. (2018). Explication and definition of mental health recovery: A systematic review. *Administration and Policy in Mental Health and Mental Health Services Research*, 45, 91–102.
- Gopal, S. & Henderson, A. (2015). Trans-cultural study of recovery from severe enduring mental illness in Chennai, India and Perth, Western Australia. *Journal of Psychosocial Rehabilitation and Mental Health*, 2, 51–57.
- Haynes, S. N., Richard, D. C. S. & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7, 238–247.
- Jacob, S., Munro, I., Taylor, B. J. & Griffiths, D. (2017). Mental health recovery: A review of the peer-reviewed published literature. *Collegian (Royal College of Nursing, Australia)*, 24, 53–61.
- de Jong, S., van Donkersgoed, R., Renard, S. et al. (2018). Social-cognitive risk factors for violence in psychosis: A discriminant function analysis. *Psychiatry Research*, 265, 93–99.
- Kempen, G. I. J. M. & van Eijk, L. M. (1995). The psychometric properties of the SSL12-I, a short scale for measuring social support in the elderly. *Social Indicators Research*, 35, 303–312.
- 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, 192–207.
- Leamy, M., Bird, V., Le Boutillier, C., Williams, J. & Slade, M. (2011). Conceptual framework for personal recovery in mental health: Systematic review and narrative synthesis. *The British Journal of Psychiatry: The Journal of Mental Science*, 199, 445–452.
- Liberman, R. P. (2012). Recovery from schizophrenia: Form follows functioning. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, 11, 161–162.
- Liberman, R. P. & Kopelowicz, A. (2005). Recovery from schizophrenia: A concept in search of research. *Psychiatric Services*, 56, 735–742.
- Mathew, S. T., Nirmala, B. P. & Kommu, J. V. S. (2018). Meaning of recovery among persons with schizophrenia: A literature review. *Journal of Psychosocial Rehabilitation and Mental Health*, 5, 73–81.
- van Nieuwenhuizen, C., Wilrycx, G., Moradi, M. & Brouwers, E. (2014). Psychometric evaluation of the Dutch version of the mental health recovery measure (MHRM). *The International Journal of Social Psychiatry*, 60, 162–168.
- Noiseux, S. & Ricard, N. (2008). Recovery as perceived by people with schizophrenia, family members and health professionals: A grounded theory. *International Journal of Nursing Studies*, 45, 1148–1162.
- Orwin, R. G. (1994). Evaluating coding decisions. In: *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.
- Roe, D., Mashlach-Eizenberg, M. & Lysaker, P. H. (2011). The relation between objective and subjective domains of recovery among persons with schizophrenia-related disorders. *Schizophrenia Research*, 131, 133–138.
- Salzer, M. S. & Brusilovskiy, E. (2014). Advancing recovery science: Reliability and validity properties of the recovery assessment scale. *Psychiatric Services*, 65, 442–453.
- Salzmann-Erikson, M. (2013). An integrative review of what contributes to personal recovery in psychiatric disabilities. *Issues in Mental Health Nursing*, 34, 185–191.
- 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, 974–980.
- Silverstein, S. M. & Bellack, A. S. (2008). A scientific agenda for the concept of recovery as it applies to schizophrenia. *Clinical Psychology Review*, 28, 1108–1124.
- 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, 1082–1095.
- Slade, M., Leamy, M., Bacon, F. et al. (2012). International differences in understanding recovery: systematic review. *Epidemiology and Psychiatric Sciences*, 21, 353–64.
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80, 99–103.
- Streiner, D. L. & Norman, G. R. (2008). *Health Measurement Scales: A Practical Guide to their Development and Use*, 4th edn. New York: Oxford University Press.

- Stuart, S. R., Tansey, L. & Quayle, E. (2017). What we talk about when we talk about recovery: A systematic review and best-fit framework synthesis of qualitative literature. *Journal of Mental Health, 26*, 291–304.
- Terwee, C. B., Bot, S. D., de Boer, M. R. *et al.* (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*, 34–42.
- Vogel, J. S., Swart, M., Slade, M., Bruins, J., van der Gaag, M. & Castelein, S. (2019). Peer support and skills training through an eating club for people with psychotic disorders: A feasibility study. *Journal of Behavior Therapy and Experimental Psychiatry, 64*, 80–86.
- van Weeghel, J., van Zelst, C., Boertien, D. & Hasson-Ohayon, I. (2019). Conceptualizations, assessments, and implications of personal recovery in mental illness: A scoping review of systematic reviews and meta-analyses. *Psychiatric Rehabilitation Journal, 42*, 169–181.
- Wood, L. & Alsawy, S. (2018). Recovery in psychosis from a service user perspective: A systematic review and thematic synthesis of current qualitative evidence. *Community Mental Health Journal, 54*, 793–804.
- World Medical Association (WMA). (2013). *Declaration of Helsinki. Amended by the 64th WMA general assembly, Fortaleza, Brazil.* Ferney-Voltaire, France: WMA Archives.
- Young, S. L. & Bullock, W. A. (2005). Mental health recovery measure (MHRM). In: T. Campbell-Orde, J. Chamberlin, J. Carpenter & H. S. Leff (Eds). *Measuring the Promise: A Compendium of Recovery Measures* (pp. 36–41). Cambridge, MA: Human Services Research Institute.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

**Data S1.** Original Subscales.

**Data S2.** Recovery Assessment Scale (Dutch version).

**Table S1.** CHIME Subscales.

**Table S2.** Internal consistency of the RAS, MHRM and NEL.