



The INSPIRE Support sub-scale: A critical view of the psychometric properties focusing on dimensionality

Jūratė Šaltytė Benth^{1,2} · Regina Skar-Fröding³ · Torleif Ruud^{1,3} · Hanne Kristin Clausen^{3,4} · Kristin Sverdvik Heiervang^{3,5}

Accepted: 24 May 2021 / Published online: 2 June 2021
© The Author(s) 2021

Abstract

The statistical methods previously used for assessing dimensionality of the INSPIRE *Support* sub-scale are often not valid, which may lead to inaccurate conclusions. The first aim of this paper was to present and discuss the methods for assessing the dimensionality of the scale. The second aim was to illustrate these methods using data from a Norwegian translation of the INSPIRE Support sub-scale. Exploratory factor analysis with several extraction and rotation methods was used to identify the dimensionality of the scale. Cronbach's alpha and omega coefficient were employed to assess the internal consistency. Using the recognized statistical approach, there was no clear structure found in the Support sub-scale, indicating that the scale might be one-dimensional. While single items or sum score of the INSPIRE Support sub-scale might cover important aspects of the personal recovery concept, there is no evidence of the five CHIME dimensions of this scale as published earlier. A more thorough assessment is called for, and results from studies using the five dimensions should be interpreted with caution.

Keywords INSPIRE · Factor analysis · Dimensionality · Internal consistency

Introduction

The INSPIRE questionnaire (Williams et al., 2015) was designed to assess service users' experiences of support from

Impact and Implications:

Even though single items and the sum score of the INSPIRE Support sub-scale cover important aspects of the personal recovery concept, there is no clear support for the five-domain structure as described in the CHIME framework. It is therefore necessary with more thorough research to gain more insight into the dimensionality of the sub-scale using recognized statistical methods.

✉ Jūratė Šaltytė Benth
jurates@medisin.uio.no

- ¹ Institute of Clinical Medicine, Campus Ahus, University of Oslo, Oslo, Norway
- ² Health Services Research Unit, Akershus University Hospital, Lørenskog, Norway
- ³ Division of Mental Health Services, Akershus University Hospital, Lørenskog, Norway
- ⁴ Norwegian National Advisory Unit on Concurrent Substance Abuse and Mental Health Disorders and Mental Health Division, Innlandet Hospital Trust, Brumunddal, Norway
- ⁵ Centre for Medical Ethics, Faculty of Medicine, University of Oslo, Oslo, Norway

health professionals in their personal recovery. The questionnaire consists of two sub-scales, Support and Relationships, and it is based on two studies containing theoretical foundations for a measure of staff support for personal recovery. In the first study, including a review of 97 publications (Leamy et al., 2011), the authors performed a systematic review and a narrative synthesis of relevant models and frameworks, resulting in five domains of the recovery process: Connectedness, Hope, Identity, Meaning and purpose, and Empowerment (CHIME). The second study (Le Bouillier et al., 2011), based on a qualitative analysis of 30 international documents, provided a theoretically justifiable classification of domains of recovery support. The INSPIRE was developed by drafting the questions, refining and reducing the initial list of questions, piloting a questionnaire, and finally psychometrically evaluating it (Williams et al., 2015). The final version of the Support sub-scale comprises the five key domains of the recovery process, as defined in CHIME, each consisting of four questions, while the Relationship sub-scale exhibits a one-dimensional structure.

Convergent validity of the questionnaire was assessed by correlating the INSPIRE sub-scales with two validated questionnaires: the Support sub-scale with the Satisfaction Index-Mental Health, a 12-item service user-rated measurement of satisfaction with services; and the Relationships sub-scale

with the Recovery-Promoting Relationship Scale, a 24-item measure of the competencies of mental health staff promoting recovery. While the Relationship sub-scale showed adequate convergent validity (correlation of 0.69), it was rather low for the Support sub-scale (correlation 0.47). The five-factor structure (CHIME) in the Support sub-scale was assessed by performing exploratory factor analysis separately in each domain, aiming at demonstrating their unidimensionality. For the Relationship sub-scale, a parallel analysis assessing whether a single underlying factor could be identified, was performed. Internal consistency of each domain of the Support sub-scale and the Relationship sub-scale was measured by Cronbach's alpha and shown to be adequate (Cronbach's alpha 0.82–0.95). Test-retest reliability was assessed by analysing changes in ratings of importance on the Support sub-scale, which were demonstrated to be small, and correlating scores of the Relationship sub-scale obtained at two time points, showing good reliability with intra-class correlation coefficient of 0.75.

The INSPIRE questionnaire has been translated into a number of languages (INSPIRE manual, 2015). However, to our knowledge only the Swedish translation of the questionnaire has been assessed psychometrically (Schön et al., 2015). The Norwegian version (available at NAPHA, 2017) has not been evaluated psychometrically. While most of the established methods for the evaluation of the psychometric properties of the questionnaires seem to work fine in the paper by Williams et al. (2015), the assessment of the dimensionality and the internal consistency of the domains of Support sub-scale of the INSPIRE questionnaire raises some concerns. While attempting to validate the Support sub-scale in the Norwegian context, a number of problems appeared, in particular when assessing the dimensionality of the sub-scale. As CHIME seems to be the most accepted framework for assessing personal recovery (van Weeghel et al., 2019), a better understanding of its domains, and studies claiming to measure these, is of importance. In this paper, the focus is on the *Support* and not the *Relationship* sub-scale.

Hence, the aim of this paper is two-folded. The first aim is to present and discuss the methods for assessing the dimensionality of the scale. For clarity, this is presented in a separate section in the manuscript (“Methodological considerations”). The second aim is to illustrate these methods by applying them on data collected by using a Norwegian translation of the INSPIRE Support sub-scale.

Methods

The INSPIRE Support sub-scale consists of 20 questions. First, the participant is asked to answer “yes” or “no” whether a certain item of the sub-scale is important for the participant's personal recovery. If the answer is “no”, the participant is

asked to proceed to the next question. In the case the item is important, the participant is asked to rate how much support the participant receives from the mental health care worker on the ordinal scale 0–4 with 0 meaning “not at all”, 1 – “not much”, 2 – “somewhat”, 3 – “quite a lot”, and 4 – “very much”.

Participants and Data

The INSPIRE questionnaire was filled in by patients as part of a patient outcome sub-study of a Norwegian research project “Implementation of Guidelines for Treatment of Psychoses” (NCT03271242). Altogether 39 community mental health centers and inpatient wards participated in the project. Inclusion criteria were: patients 16+ years old and having a diagnosis from the F20–29 section in the International Statistical Classification of Diseases and Related Health Problems (ICD-10) (World Health Organization, 1992). There were no exclusion criteria except that patients had to be able to understand and answer the questionnaires in Norwegian. Patients were included at the level of the mental health clinics of six health trusts, and questionnaires were administered to the patients by the clinicians. Inclusion of patients started in the beginning of June 2016 and was completed in March 2017. Only baseline data were used in this study, and in total, 325 patients with mean age of 40 (SD 12.7) years, 41% female and 21% with higher education, were included. Among the patients, 54% had a diagnosis of schizophrenia, 20% had a diagnosis of schizoaffective disorder, and 27% had another diagnosis from the F20–29 section in ICD-10. Further patient characteristics and study details have been reported earlier (Skar-Frøding et al., 2021).

Methodological Considerations

When developing a new questionnaire, the presumed dimensionality is typically assessed by exploratory factor analysis. In general, the items belonging to one dimension are assumed to reflect the same underlying construct. After the dimensions are extracted, the internal consistency for each dimension (reliability across the items) is evaluated. These issues are elaborated on next.

The exploratory factor analysis is a tool for analysing the correlations among a large number of items with the purpose of grouping those items into constructs (or factors in the statistical language). Items within each factor are supposed to be highly correlated, while items from different factors should not be correlated in any substantial way in order to conclude with multiple dimensions. In other words, exploratory factor analysis identifies the necessary amount of dimensions representing data. According to Armor (1973-1974), the

correlation analysis and factor analysis should be the first step in scale development, as this is a tool for determining the dimensionality of the scale. If factor analysis generates multiple factors, the multidimensionality of the scale is confirmed, and the correlations among items contributing to different factors are nearly zero. A multidimensional measure thus consists of a number of unidimensional measures (Whidhiarso & Ravand, 2014).

However, it is problematic to conclude that the measure is multidimensional from multiple unidimensional solutions without performing factor analysis on all items simultaneously. Performing factor analysis on each dimension separately with the aim to demonstrate its unidimensionality is not enough, as it cannot identify items belonging to different dimensions, but still overlapping in a substantial degree according to factor loadings (i.e. high cross-loadings). According to Costello and Osborne (2005), “a ‘crossloading’ item is an item that loads at 0.32 or higher on two or more factors”. Ziegler and Hagemann (2015) emphasize that when constructing the scale, it should at least be ensured that the items in one construct do not load on other constructs. Crutzen and Peters (2017) also call for factor-analytic evidence before assessing internal structure of the scale.

When the dimensionality of the scale is established, the internal consistency can be assessed. One of the most popular measures of internal consistency of a scale is Cronbach’s alpha (Cronbach, 1951). A value of the alpha close to 1 indicates a good internal consistency of the scale. Since Cronbach’s alpha is a function of the number of items, increasing number of items may increase the alpha as well. Cortina (1993) has shown that for scales with more than 14 items, Cronbach’s alpha may exceed 0.70 even in the case of two orthogonal dimensions, and be even higher if the dimensions are correlated, which often is the case in practice. Moreover, if the covariances among the items are very high, the resulting alpha will also be high, often above 0.90, which might indicate redundancy of some items in the scale, and consequently, a high value is not necessarily a sign of a consistent scale (Streiner, 2003).

The important feature of the alpha is that it is not a test of unidimensionality (Streiner, 2003). It rather assumes that the sample of items is unidimensional and may cause underestimated reliability if this assumption is violated (Tavakol & Dennick, 2011). Even if other assumptions for alpha to be a valid measure hold, high values may appear in the scales with multiple dimensions. Thus, alpha can only be used to confirm unidimensionality, not assess it (Cortina, 1993; Green et al., 1977). If the factor analysis results in more than one dimension, Cronbach’s alpha should be calculated within each dimension separately (Whidhiarso & Ravand, 2014).

The alpha is based on a so-called assumption of *essential tau-equivalence*, requiring that all items have the same

variances, that covariances between items are equal, that items measure the same underlying construct, and have equal factor loadings (Trizano-Hermosilla & Alvarado, 2016). If the assumption of essential tau-equivalence is violated, the scale reliability may be underestimated, and the alpha can only be used as a lower bound approximation of the scale reliability (Cronbach, 1951).

In practice, a *congeneric* model only requiring the items to measure one underlying construct (i.e. unidimensionality) is more realistic. In the case of a congeneric model, omega reliability coefficient is a better measure of scale consistency (McDonald, 1999). Omega coefficient is suitable both when essential tau-equivalence exists and when the items constitute congeneric measurement (Trizano-Hermosilla & Alvarado, 2016). Omega coefficient applied to essentially tau-equivalent measurements (single factor model with equal loadings) would be equal to the alpha. For congeneric measurements, its value will be higher than the alpha (Whidhiarso & Ravand, 2014).

Statistical Analyses

Frequencies and percentages of “Not important” answers and means and standard deviations (SDs) for the level of support were presented, and the number of missing values for each item was reported (Table 1).

Exploratory factor analysis was performed to assess the dimensionality of the Norwegian translation of the INSPIRE Support sub-scale. The exploratory, not the confirmatory factor analysis was chosen, because the sub-scale has not previously been psychometrically assessed in a Norwegian context. Another and most important reason for using exploratory factor analysis was that an approach used by Williams et al. (2015) to assess the dimensionality of the sub-scale is not valid, and hence there is yet no factor structure to be confirmed.

In the factor analysis, cases with at least one missing value on the included items are eliminated. List-wise elimination of cases with missing items resulted in substantially reduced sample size, and thus power. The exploratory factor analysis was therefore performed on a matrix of polychoric correlations among the items with the correlations calculated on all available pairs of observations. Polychoric instead of Pearson correlations are preferred in factor analysis of ordinal variables, since Pearson correlations tend to be underestimated, which again results in smaller values of factor loadings (Holgado-Tello et al., 2010).

Iterated principal factors, principal factors and principal-component factors extraction methods, and *varimax* and *oblique promax* rotation methods were applied. The number of factors was assessed in several ways. Kaiser’s criterion of eigenvalue >1 in combination with the scree-test followed by

Table 1 Descriptive statistics for ratings of the INSPIRE Support sub-scale, $N = 325$

Item	Not important, N (%)	Important, N (%)	Number of missing (%)	Support, Mean (SD)
1. Feeling supported by other people	37 (11.4)	284 (87.4)	4 (1.2)	2.8 (0.9)
2. Having positive relationships with other people	30 (9.2)	290 (89.2)	5 (1.5)	2.7 (0.9)
3. Having support from other people who use services	98 (30.2)	217 (66.8)	10 (3.1)	2.6 (1.0)
4. Feeling part of my community	74 (22.8)	246 (75.7)	5 (1.5)	2.6 (1.0)
5. Feeling hopeful about my future	58 (17.8)	259 (79.7)	8 (2.5)	2.6 (1.0)
6. Believing that I can recover	40 (12.3)	280 (86.2)	5 (1.5)	2.8 (1.0)
7. Feeling motivated to make changes	68 (20.9)	250 (76.9)	7 (2.2)	2.6 (1.0)
8. Having hopes and dreams for the future	51 (15.7)	268 (82.5)	6 (1.8)	2.6 (1.1)
9. Feeling I can deal with stigma	91 (28.0)	214 (65.8)	20 (6.2)	2.3 (1.1)
10. Feeling good about myself	70 (21.5)	247 (76.0)	8 (2.5)	2.6 (1.0)
11. Having my spiritual beliefs respected	99 (30.5)	215 (66.2)	11 (3.4)	2.6 (1.0)
12. Having my ethical/cultural/racial identity respected	105 (32.3)	207 (63.7)	13 (4.0)	2.7 (1.1)
13. Understanding my mental health experience	63 (19.4)	251 (77.2)	11 (3.4)	2.7 (1.0)
14. Doing things that mean something to me	34 (10.5)	285 (87.7)	6 (1.8)	2.7 (0.9)
15. Rebuilding my life after difficult experiences	52 (16.0)	264 (81.2)	9 (2.8)	2.8 (1.0)
16. Having a good quality of life	40 (12.3)	278 (85.5)	7 (2.2)	2.7 (1.0)
17. Feeling in control of my life	55 (16.9)	262 (80.6)	8 (2.5)	2.7 (1.0)
18. Being able to manage my mental health	37 (11.4)	281 (86.5)	7 (2.2)	2.8 (0.9)
19. Trying new things	104 (32.0)	216 (66.5)	5 (1.5)	2.5 (1.0)
20. Building on my strength	46 (14.2)	271 (83.4)	8 (2.5)	2.6 (1.0)

parallel analysis was applied first. Next, to partially mimic the analyses by Williams et al. (2015), the five-factor solution was requested. Finally, entirely for comparison purposes, a factor analysis was performed separately for each dimension anticipated by Williams et al. (2015).

Other approaches for dealing with missing values were considered as well. Keeping in mind the algorithm for calculating the sum score for the Support sub-scale (INSPIRE, 2015), an imputation of missing values with mean of existing items for each patient was performed, although not recommended (Eekhaut et al., 2014). The approach using maximum likelihood with the expectation-maximization algorithm to estimate the covariance matrix (*mi estimate* function in the STATA) suggested by Graham (2009) was employed as well.

To assess the internal consistency of the identified dimensions in the Support sub-scale, the Cronbach's alpha and the omega coefficient were calculated.

The statistical analyses were performed with STATA v16 and SPSS v26.0.

Results

Description of the Support Sub-Scale

The number of missing answers to whether a certain item of the Support sub-scale is important varied between 4 and 20 (Table 1). The item "Having positive relationships with other people" was most often chosen as being important by $N = 290$ (89.2%) of the 325 participants. The item reported as being the

least important was "Having my ethical/cultural/racial identity respected" ($N = 207$, 63.7%).

Among participants for whom support is important, the mean ratings varied between 2.3 and 2.8 with SDs between 0.9 and 1.1 for all items (Table 1). Due to some participants answering "Not important" or not answering at all to whether a certain item is important the number of missing rating values varied between 35 and 118. As a result, there were only 66 (20.3%) participants of 325 with no missing values on all 20 ratings.

Exploratory Factor Analysis

The polychoric correlations between pairs of items among 325 participants are presented in Table 2. If the INSPIRE Support sub-scale clearly would comprise five dimensions, one would expect strong correlations along the diagonal in 4×4 -blocks and substantially weaker elsewhere. However, no clear pattern can be seen in the correlation matrix in Table 2. Nevertheless, the matrix of polychoric correlations was used as input in an exploratory factor analysis.

According to Kaiser's criterion of eigenvalue >1 combined with scree-plot followed by parallel analysis, the 20 items comprised a three-factor solution. The factor structure was stable despite of the extraction method used and the rotation applied. Therefore, only the results of the analysis based on the iterated principal factors extraction method with *varimax* rotation were presented (Table 3, columns 1–3 under three-factor solution). Among the three factors identified, the largest eigenvalue of 11.5 attributed to the first factor, while the second and third factors contributed little (eigenvalues of 1.1 and

0.9, respectively). The three factors explained 66.2%, 6.3% and 5.0% of the total variance, respectively, providing no strong evidence of more than one dimension. The factors contained items that were randomly distributed with respect to the expected dimensions, and there were many strong cross-loadings (above 0.32) present, indicating no structure similar to CHIME in the Support sub-scale. When a five-factor structure was required, the exploratory factor analysis suggested similar conclusion (Table 3, columns 1–5 under five-factor structure).

Other Ways of Handling Missing Values

Exploratory factor analysis performed on a data set where missing values were imputed with mean of existing items for each patient produced a one-factor solution. In the case where the covariance matrix was estimated by expectation-maximization method, also one-factor solution was obtained implying no dimensions in the Support sub-scale. Results for these two analyses are not presented, but available upon request.

Assessing Internal Consistency

Even though SDs were quite equal for all 20 items, the means varied to some extent. Also the correlations among the items and the factor loadings in all factor solutions were rather

different. This indicates violation of the assumption of essentially tau-equivalent measurement and hence makes Cronbach's alpha an inadequate measure for internal consistency. Therefore, omega coefficient should be prioritised. However, we also report Cronbach's alpha to make our results comparable to Williams et al. (2015).

According to Cronbach's alpha and the omega coefficient for three- and five-factor solutions presented in Table 3, the internal consistency is good in all dimensions and in both solutions, however, it seems that Cronbach's alpha underestimates the internal consistency in the identified dimensions. Many items loading on factor 2 and 3 in the three-factor solution load together in the five-factor solution. Most of the items in factor 1 in three-factor solution remain in factor 1 in five-factor solution as well, and only five items load on factor 2, also suggesting a one-factor solution.

For comparison, we also assessed the internal consistency in the same way as Williams et al. (2015), i.e. we assumed the presence of five domains, and performed exploratory factor analysis and calculated Cronbach's alpha within those domains. Five unidimensional solutions were obtained and the alpha coefficients for each of the five domains were, correspondingly, 0.79, 0.87, 0.79, 0.85, and 0.86. As the assumption of essentially tau-equivalent model is likely violated, the omega coefficients were calculated and found to be 0.85, 0.90, 0.84, 0.89, and 0.88, respectively. Both the alpha and omega coefficients showed good internal consistency.

Table 2 Polychoric correlations among the INSPIRE Support sub-scale items calculated between pairs of items using all available observations. Item numbering is the same as in Table 1

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		0.63	0.65	0.44	0.44	0.43	0.49	0.47	0.44	0.47	0.46	0.40	0.50	0.48	0.50	0.47	0.42	0.46	0.46	0.41
2			0.63	0.56	0.54	0.53	0.50	0.59	0.42	0.51	0.43	0.36	0.56	0.53	0.50	0.53	0.47	0.49	0.49	0.50
3				0.53	0.46	0.41	0.47	0.48	0.35	0.51	0.47	0.51	0.55	0.55	0.52	0.49	0.50	0.44	0.53	0.45
4					0.59	0.58	0.52	0.63	0.42	0.53	0.50	0.50	0.58	0.57	0.60	0.63	0.55	0.59	0.60	0.62
5						0.74	0.64	0.72	0.43	0.69	0.46	0.43	0.57	0.55	0.66	0.67	0.64	0.60	0.54	0.57
6							0.69	0.72	0.48	0.56	0.47	0.38	0.56	0.54	0.66	0.70	0.64	0.62	0.63	0.56
7								0.66	0.56	0.59	0.56	0.49	0.56	0.55	0.67	0.57	0.60	0.61	0.61	0.63
8									0.57	0.59	0.55	0.43	0.56	0.67	0.65	0.67	0.71	0.66	0.66	0.69
9										0.59	0.57	0.52	0.54	0.43	0.58	0.44	0.44	0.51	0.57	0.53
10											0.51	0.44	0.55	0.54	0.64	0.62	0.62	0.56	0.50	0.59
11												0.72	0.54	0.45	0.61	0.49	0.47	0.52	0.65	0.58
12													0.43	0.45	0.55	0.48	0.48	0.34	0.48	0.48
13														0.66	0.67	0.54	0.54	0.65	0.56	0.63
14															0.70	0.66	0.55	0.63	0.53	0.72
15																0.74	0.67	0.77	0.62	0.70
16																	0.75	0.71	0.59	0.64
17																		0.73	0.65	0.60
18																			0.60	0.70
19																				0.63
20																				

Table 3 Results of exploratory factor analysis; iterated principal factors extraction method with varimax rotation. Strongest loadings for each item are highlighted by bold face, while italic highlights high cross-loadings (loadings of 0.32 or higher (Costello & Osborne, 2005))

Item number	Three-factor solution			Five-factor solution				
	1	2	3	1	2	3	4	5
1. Feeling supported by other people	0.27	0.26	0.67	0.22	0.18	0.65	0.19	0.20
2. Having positive relationships with other people	<i>0.40</i>	0.16	0.68	<i>0.34</i>	0.22	0.66	0.11	0.17
3. Having support from other people who use services	0.25	0.29	0.75	0.22	0.22	0.71	0.30	0.02
4. Feeling part of my community	0.55	0.31	<i>0.39</i>	0.44	<i>0.38</i>	<i>0.38</i>	0.29	0.08
5. Feeling hopeful about my future	0.72	0.20	0.30	0.73	0.21	0.31	0.15	0.18
6. Believing that I can recover	0.75	0.22	0.24	0.73	0.24	0.24	0.14	0.24
7. Feeling motivated to make changes	0.61	<i>0.40</i>	0.28	0.53	0.28	0.28	0.27	<i>0.36</i>
8. Having hopes and dreams for the future	0.74	0.29	0.30	0.65	<i>0.36</i>	0.29	0.19	0.27
9. Feeling I can deal with stigma	<i>0.41</i>	0.53	0.21	0.27	0.23	0.20	<i>0.36</i>	0.60
10. Feeling good about myself	0.58	<i>0.33</i>	<i>0.33</i>	0.52	0.25	<i>0.34</i>	0.23	0.30
11. Having my spiritual beliefs respected	0.31	0.81	0.24	0.23	0.25	0.24	0.67	<i>0.35</i>
12. Having my ethical/cultural/racial identity respected	0.24	0.69	0.27	0.21	0.16	0.25	0.79	0.11
13. Understanding my mental health experience	0.52	<i>0.34</i>	<i>0.43</i>	0.28	0.51	<i>0.43</i>	0.20	0.30
14. Doing things that mean something to me	0.60	0.26	<i>0.40</i>	<i>0.33</i>	0.62	<i>0.39</i>	0.18	0.10
15. Rebuilding my life after difficult experiences	0.69	<i>0.43</i>	0.28	<i>0.47</i>	0.55	0.26	<i>0.33</i>	0.23
16. Having a good quality of life	0.74	0.24	0.29	0.64	<i>0.45</i>	0.27	0.26	0.00
17. Feeling in control of my life	0.72	0.27	0.25	0.67	<i>0.40</i>	0.20	0.29	0.02
18. Being able to manage my mental health	0.75	0.26	0.24	<i>0.50</i>	0.64	0.20	0.13	0.22
19. Trying new things	0.55	<i>0.47</i>	0.30	0.44	<i>0.34</i>	0.27	<i>0.38</i>	0.30
20. Building on my strength	0.66	<i>0.40</i>	0.25	<i>0.38</i>	0.61	0.22	0.27	0.25
Cronbach's alpha ^a	0.92	0.77	0.79	0.93	0.89	0.79	0.79	
Omega ^b	0.95	0.88	0.88	0.93	0.89	0.88	0.84	0.75

^a Cronbach's alpha is calculated for items with loadings in bold face; ^b Omega is calculated by including factor loadings on all items

Interestingly, Cronbach's alphas calculated for *all possible* ($n = 4845$) combinations of four items are all similar (Fig. 1) and indicate quite good internal consistency with min 0.70 and max 0.88, and average alpha 0.79 with 95% CI of (0.791; 0.793). Also the alpha coefficient for all 20 items together was estimated to be 0.95 (while the omega was 0.96).

Discussion

The first step in assessing dimensionality of a scale should be a correlation analysis followed by a factor analysis performed on all items. Only after the dimensionality is established, the internal consistency can be assessed. The measures of internal consistency do not measure the dimensionality, they only confirm it. In other words, internal consistency is a necessary but not sufficient condition for measuring unidimensionality in a sample of items (Green et al., 1977). In the case of multiple dimensions, internal consistency should be assessed for each dimension separately as the statistics used for this purpose rely on the assumption that items measure one underlying construct (unidimensionality). The alpha coefficient for internal consistency can only be used when the items comprise a tau-equivalent measure, an assumption which is often violated. In practice, congeneric models are more realistic and other statistics, for example the omega coefficient, for internal consistency should be preferred.

Williams et al. (2015) assessed the dimensionality of the INSPIRE questionnaire by performing factor analysis for each *intended* dimension separately, arguing with small sample size, and in such way assuming that certain items measure *only* the anticipated dimension. Even though their five exploratory factor analyses resulted in five unidimensional solutions, there is no way to quantify the cross-loadings among the items. The correlation matrix could give some insight into cross-correlations among items, but it is not presented in the paper by Williams et al. (2015). Schön et al. (2015) skip factor analysis completely and assume the five-factor structure proposed by Williams et al. (2015) with no statistical evidence.

The correlations in the present study showed no clear pattern, not regarding the five CHIME domains introduced by Williams et al. (2015), nor any other interpretable structure in the Support sub-scale. A number of factor analyses performed resulted in one dimension differing substantially from the proposed five-factor structure. This calls for more research assessing the dimensionality of the INSPIRE Support sub-scale on different data sets.

Both Williams et al. (2015) and Schön et al. (2015) assessed the internal consistency of each CHIME domain using Cronbach's alpha. However, the essential tau-equivalence assumption was not tested or discussed in either paper. The means of the 20 items varied quite a bit in both studies. As no correlations among the items were present, it

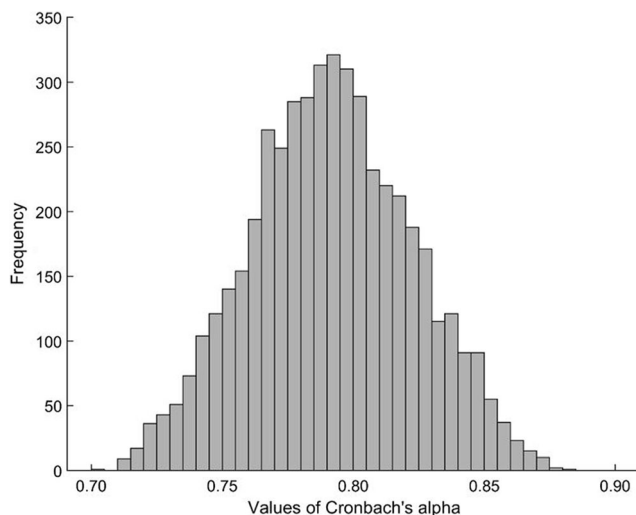


Fig. 1 Distribution of Cronbach's alphas calculated for all possible combinations of 4 out of 20 items, $N = 4845$

was difficult to assess the assumption further. It seems like the omega coefficient, which is suitable for both essentially tau-equivalent and congeneric models, would be a more appropriate statistic for internal consistency.

In the present study the assumption of essential tau-equivalence was likely not met, and the omega coefficient was therefore estimated. However, for comparison purposes also Cronbach's alpha was calculated. The omega coefficients and Cronbach's alphas differed slightly, indicating that the items do not comprise an essential tau-equivalent measure. A good internal consistency was demonstrated independently of the number of domains as well as for all 20 items assessed as one domain. As expected, both the alpha and the omega coefficients indicated redundancy of some items in later case. The alpha coefficient calculated for the five domains suggested by Williams et al. (2015) and "verified" by Schön et al. (2015), also showed good internal consistency in the present study. For illustrational purposes, the alpha coefficient was calculated for all the possible 4845 combinations of four items out of 20. The variation in these calculations was strikingly small with average alpha of 0.79 and 95% CI of (0.791; 0.793), also indicating rather good internal consistency of most of the 4845 possible "domains".

Another important aspect not discussed by Williams et al. (2015) is an issue related to a number of missing ratings on the Support sub-scale, which occur every time the person answers "no" on the question whether a certain item of the sub-scale is important for the participant's personal recovery. Missing rating may also appear if the person is not rating a certain item even though the item is important for his or her personal recovery. According to the INSPIRE Support sub-scale scoring manual (INSPIRE, 2015), the sum score cannot be calculated only if all 20 questions are rated "no" or left blank. It can however be calculated for persons rating all 20 items or, for example, only one item, implying that the missing values are

imputed with the persons mean rating. One can discuss a quality of a sum score if only a few items are rated, but this is out of the scope of this paper. However, the way the sum score is calculated, may indicate that a person's mean rating could have been used for imputing missing values on relevant items in previous studies. As the factor analysis struggles with cases with at least one missing value, this may seem an appealing solution to the problem but the mean imputation is not a recommended practice in the context of factor analysis as it reduces the variation remarkably Eakhaut et al. (2014). It is unfortunately unclear how Williams et al. (2015) dealt with missing ratings in their publication.

The number of missing ratings was quite high in the present study and excluding patients with at least one missing value among 20 items would result in a substantially reduced sample size. Therefore several approaches for dealing with missing values were considered. The main results of this study are based on an exploratory factor analysis performed on polychoric correlations, the approach also used by Williams et al. (2015). The polychoric correlations were chosen because the ratings constitute an ordinal scale. But another – just as important – reason for using correlation matrix as input to exploratory factor analysis was the handling of missing values. The correlations were calculated between all pairs of items, including all possible observations. Two other approaches handling a large amount of missing values were applied – imputation with a person's mean value as well as estimating the covariance matrix by the expectation maximization method. Both resulted in one-factor solution with very similar loadings.

There is no clear support for the five-domain structure described in the CHIME framework by Williams et al. (2015). The analysis in the present study suggests one single dimension. There is therefore a clear need for more research through new and larger studies to gain more insight into the dimensionality of the Support sub-scale using the recognized statistical methods. However, single items and the sum score of the questionnaire still cover important aspects of the personal recovery concept. Previous research has shown that a large sample of service users with psychosis rates a clear majority of the items in the INSPIRE Support sub-scale as important for their personal recovery process (Skar-Fröding et al., 2021). Hence, the questionnaire can still be used to investigate what personal recovery areas are of importance and whether service users experience that they receive support for this. Future research could investigate the important knowledge gap of how personal recovery processes take place and how support for this is perceived in patients with other common mental disorders such as mood disorders or anxiety.

Acknowledgments The study funder, the South-Eastern Regional Health Authority, will not gain or lose financially from the publication of this

manuscript, either now or in the future. The views expressed are those of the authors and not necessarily those of South-Eastern Regional Health Authority. We thank the six participating health trust and their service users for participating in the study.

Funding Open access funding provided by Akershus University Hospital (AHUS). The study was funded by the South-Eastern Norway Regional Health Authority (grant number 2015106). Only patients giving written informed consent were included into study. The study was approved by the Regional Committee for Ethics in Research (REK Sørøst B 2015/2169). According to written consent from the participants, distribution of the data file to others than the research group that conducted the study is not allowed. Researchers wanting to obtain access to the data may contact the corresponding author, who will forward the request to the principal investigator.

Data Availability According to written consent from the participants, distribution of the data file to others than the research group that conducted the study is not allowed. Researchers wanting to obtain access to the data may contact the corresponding author, who will forward the request to the principal investigator.

Declarations

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

We have no known conflicts of interest to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Armor, D. J. (1973-1974). Theta reliability and factor scaling. *Sociological Methodology*, 5, 17–50 <http://www.jstor.org/stable/270831?origin=JSTOR-pdf>.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104 <https://psycnet.apa.org/doi/10.1037/0021-9010.78.1.98>.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, 10(Article 7). <https://scholarworks.umass.edu/pare/vol10/iss1/7>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of test. *Psychometrika*, 16, 297–334. <https://doi.org/10.1007/BF02310555>.
- Crutzen, R., & Peters, G.-J. Y. (2017). Scale quality: Alpha is an inadequate estimate and factor analytic evidence is needed first of all. *Health Psychology Review*, 11(3), 242–247. <https://doi.org/10.1080/17437199.2015.1124240>.
- Eekhout, I., de Vet, H. C. W., Twisk, J. W. R., Brand, J. P. L., de Boer, M. R., & Heymans, M. W. (2014). Missing data in multi-item instrument were best handled by multiple imputation at the item score level. *Journal of Clinical Epidemiology*, 67(3), 335–342. <https://doi.org/10.1016/j.jclinepi.2013.09.009>.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>.
- Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37, 827–838. <https://doi.org/10.1177/001316447703700403>.
- Holgado-Tello, F. P., Chacon-Moscoso, S., Barbero-Garcia, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, 44, 153–166. <https://doi.org/10.1007/s1135-008-9190-y>.
- INSPIRE – Research into recovery and wellbeing (2015). Retrieved October 15, 2018, from <http://www.researchintorecovery.com/INSPIRE#s9>.
- Le Boutillier, C., Leamy, M., Bird, V. J., Davidson, L., Williams, J., & Slade, M. (2011). What does recovery mean in practice? A qualitative analysis of international recovery-oriented practice guidance. *Psychiatric Services*, 62(12), 1470–1476. <https://doi.org/10.1176/appi.ps.001312011>.
- 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(6), 445–452. <https://doi.org/10.1192/bjp.bp.110.083733>.
- McDonald, R. (1999). Test theory: A unified treatment. *Lawrence Erlbaum Associates/Mahwah, New Jersey*. <https://doi.org/10.4324/9781410601087>.
- Nasjonalt kompetansesenter for psykisk helsearbeid (2017). Retrieved November 1, 2018 from <https://www.napha.no/content/20687/INSPIRE-pa-norsk>
- Schön, U.-K., Svedberg, P., & Rosenberg, D. (2015). Evaluating the INSPIRE measure of staff support for personal recovery in a Swedish psychiatric context. *Nordic Journal of Psychiatry*, 69(4), 275–281. <https://doi.org/10.3109/08039488.2014.972453>.
- Skar-Frøding, R., Clausen, H. K., Benth, J. Š., Ruud, T., Slade, M., & Heiervang, K. S. (2021). The importance of personal recovery and perceived recovery support amongst service users with psychosis. *Psychiatric Services*. <https://doi.org/10.1176/appi.ps.202000223>
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personal Assessment*, 80(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>.
- Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measures. *Frontiers in Psychology*, www.frontiersin.org/articles/10.3389/fpsyg.2016.00769/full
- 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(2), 169–181. <https://doi.org/10.1037/prj0000356>.
- Whidharso, W., & Ravand, H. (2014). Estimating reliability for multidimensional measures: A pedagogical illustration. *Review of Psychology*, 21(2), 111–121.

- Williams, J., Leamy, M., Bird, V., Le Boutillier, C., Norton, S., Pesola, F., & Slade, M. (2015). Development and evaluation of the INSPIRE measure of staff support for personal recovery. *Social Psychiatry and Psychiatric Epidemiology*, *50*(5), 777–786. <https://doi.org/10.1007/s00127-014-0983-0>.
- World Health Organization: The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. Geneva, World Health Organization (1992). <https://apps.who.int/iris/handle/10665/37958>
- Ziegler, M., & Hagemann, D. (2015). Testing the unidimensionality of items. *European Journal of Psychological Assessment*, *31*(4), 231–237. <https://doi.org/10.1027/1015-5759/a000309>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.