Feasibility, validity and reliability of the Dutch translation of INCRESE (INCRESE-NL) inventory to characterize mental health recovery narratives

Mitch N. Lases* [1,2]
Jojanneke Bruins [1,3]
Fiona Ng[4]
Stefan Rennick-Egglestone[4]
Joy Llewellyn-Beardsley[4]
Mike Slade[4,5]
Nienke van Sambeek[6]
Floortje Scheepers[6]
Ingrid D.C. van Balkom[3,7]
Stynke Castelein [1,2,3]

[1] Lentis Psychiatric Institute, Lentis Research, the Netherlands

[2] University of Groningen, Faculty of Behavioural and Social Sciences, the Netherlands

[3] University of Groningen, University Medical Centre Groningen, the Netherlands

[4] University of Nottingham, United Kingdom

[5] Nord University, Faculty of Nursing and Health Sciences, Norway.

[6] University Medical Center Utrecht, the Netherlands

[7] Autism Team Northern-Netherlands, Jonx Department of (Youth) Mental Health and Autism, Lentis Psychiatric Institute, the Netherlands

Keywords: Mental Health, Stories, Recovery, Narratives, User, Experiences, Intercoder reliability

This article has been peer reviewed.
Corresponding author: Mitch Lases
Email: m.n.lases@rug.nl

Abstract

Objective: Methods for characterizing mental health recovery narratives are needed to enable the selection of helpful narratives which can support the recovery processes of other people. The Inventory of Characteristics of Recovery Stories (INCRESE) is an existing English-language tool to characterize recovery narratives, which is not available in Dutch. The aims of this study were to create a Dutch translation of INCRESE and to evaluate the feasibility of sustained and routine use, as well as its validity and reliability.
Research design and methods: INCRESE was translated into Dutch (INCRESE-NL) using an established translation methodology. Six coders with different professional backgrounds rated 30 purposively selected Dutch narratives using INCRESE-NL and then completed an evaluation survey. Feasibility was assessed qualitatively and content validity both qualitatively and quantitatively. Gwet's AC1 agreement coefficient was used to calculate intercoder reliability.
Results: Coders evaluated INCRESE-NL as a suitable instrument to capture important characteristics of recovery narratives with an acceptable administrative burden. The content validity index was sufficient for 5/7 sections of INCRESE-NL. Most items were found to be clear and suitable for capturing the corresponding characteristic, although some missed strict coding rules or clear descriptions, especially in section 4 on narrative characteristics. Gwet's AC1 could be calculated for 67/77 items, of which 62 scored sufficient and five scored below threshold. Gwet's AC1 coefficients corresponded well with the original INCRESE Fleiss Kappa values.
Conclusions: INCRESE-NL has sufficient feasibility, validity, and reliability and can be used to characterize Dutch recovery narratives.

Background

Recovery narratives of people with lived experiences of mental health issues can offer important perspectives regarding the multifaceted recovery process. Recovery is unique to an individual, unfolds over time, occurs in its own time, and is not intrinsically definite. It is about living a meaningful life, regardless of any remaining (mental) health issues.[1, 2]

The use of recovery narratives is increasingly present within and outside of mental health services, for example through publicly available websites with collections of recovery narratives.[3] A recent systematic review about uses and misuses of recorded recovery narratives in healthcare and community settings distinguished five different application levels.[4] The first application is on a *political level*, for example, when storytelling is used by policy makers of a political party to obtain votes. The second level is on a *societal level*, where personal experience accounts are being used, for example, as part of anti-stigma campaigns or to help people who (may want to) seek mental health treatment. On a *community level,* recovery narratives can be used to connect people with different perspectives or to promote fundraising. On a *service level,* mental health-oriented experiences can be used to evaluate treatment and to develop new clinical theories and practices. Last, on an *individual level,* stories of lived experience can be used for self-advocacy of the narrator or as a therapeutic tool in an intervention.

Besides useful applications, there is also a risk that using narratives may have negative consequences.[4] One example is that a *narrative* that is published as part of a research study may end up in another location, such as a magazine, without the author's consent. Secondly, prior to discharge, a *narrator* may feel they should or even may be forced to positively review their care organization, even if the experience may not have been positive at all. The third example is related to the *audience*, who may misinterpret narrator intention or even be triggered into self-harm.[4] For example, reading the narratives of other individuals with eating disorders[5] or self-harming behaviour[6] may increase normalization, identification, competition, and imitation.

An interview study has indicated that narratives may have positive or negative effects on recipients with mental health problems.[7] Helpful outcomes that were derived from qualitative analyses were connectedness, validation, hope, appreciation, empowerment, reference shifting, and stigma reduction. However, lived experience narratives can also be harmful and lead to feelings of inadequacy, less connectedness, and pessimism about the future, and they may add to the distress of the recipients. An essential factor necessary to create an impact, either positive or negative, is that the recipient perceives a connection with the narrator or narrative content.[7] Narratives can also make an impact through a recipient learning something or feeling empathy for the narrator.

Mental health recovery narratives portraying a range of narrator's characteristics may maximize the possibility of positive impact on perceived connection by recipients.[8] Another study showed that female participants and those from a younger age group generally experienced more connection with narrators, and that matching on ethnicity may have beneficial impact.[9] Connectedness was also greater when the narrator was not fully recovered but still experiencing difficulties while living a meaningful life

nonetheless.[9] It helped recipients to realize that recovery and living a meaningful life are achievable goals. Narratives that were perceived as authentic by the recipient also increased connection. Being engaged by a story through perceived connection made the messages more convincing to the recipient.

To be able to use recovery narratives actively in mental health care settings, it is important to have information about their content to help predict and actively create an impact. Capturing the key features of a recovery narrative can help to categorize them for targeted use by mental health services, patients, and researchers. A good match might increase the chances of experiencing connectedness and the desired impact for recipients. Characterization of narratives can also help to warn people about potential harmful content and to help individuals avoid those narratives that may have a negative impact on them. Awareness of important details of narratives can help to promote beneficial impact and to find relations between contextual features and the origin of mental health issues and recovery.

A conceptual framework on important characteristics of mental health recovery narratives has recently been developed in the United Kingdom.[10] This led to the development of the English-language Inventory of Characteristics of Recovery Stories (INCRESE).[11] INCRESE makes it possible to capture the most relevant elements of narratives. To date, there is no instrument available in the Dutch language to characterize recovery narratives. Considering the good psychometric properties of INCRESE, we hypothesize that this instrument is also suitable to capture important information from Dutch recovery narratives, produced by narrators from the Netherlands. Our first aim in this study is thus to create a Dutch translation of INCRESE (INCRESE-NL) and to adapt this translation to cultural relevance. The second aim is to examine the translated instrument in relation to feasibility, validity, and reliability.

Methods

Description of INCRESE

INCRESE contains 77 items categorized into seven sections, which capture the most relevant elements of narratives.[11] The instrument allows for coding descriptive features of the narrative (i.e., narrative mode and length/duration), narrator (i.e. , demographics and illness-specific information) and characteristics, such as genre (e.g., escape or endurance), positioning (e.g., recovery within or outside services), tone (e.g., generally upbeat or generally negative), and relationship with recovery (e.g., living well or surviving day-to-day). There are also separate sections for content warnings and turning points. INCRESE concludes with 29 items on content which may be present in the narrative (e.g., pregnancy/birth, family, and stigma). Feasibility, acceptability, and reliability of the original INCRESE have been positively evaluated. The original coders reported the instrument as feasible having an acceptable administrative burden, clear instructions, and descriptions.[11] Most items were evaluated to be straightforward to answer but with some items in Section Four (Narrative Characteristics) as more subjective. INCRESE was acceptable to coders and people with lived experience of mental health problems and further improved by refinements following the pilot evaluation. Coders experienced positive feelings of inspiration during the coding process but also found it hard to read about the

difficulties and trauma of some of the narrators. Subsequently, wellbeing information was added to the instrument's instructions, and strategies for self-care and dealing with negative effects were offered to the coders. The intercoder and test-retest reliability of the original INCRESE ranged from moderate to perfect agreement.

Study design

Part A: Translation phase

In the translation phase, the main aim was to translate the original INCRESE to create INCRESE-NL. We followed the updated practices of the consortium translation process for patient reported outcome (PRO) instruments.[12] The process of creating a Dutch translation involved 11 steps, including a forward-backward translation, semantic corrections, proofreading, and testing the instrument. The complete translation process is shown in *Table 1*. During the proofreading step, four project team members (including colleagues with lived experience of mental health problems and recovery) checked for language errors, layout, and cultural relevance of items but also gave their overall impression of INCRESE-NL. Adaptations were only made if the conceptual equivalence and consistency with the original INCRESE remained intact.

Table 1 *INCRESE to INCRESE-NL: Translation and validation process steps*

Step number	Step name	Translation approach used in INCRESE-NL (involved authors: ML, JB1, MR, SC, IvB, JB2, MR)
1	Preparation	Obtaining translation permission from developer, deciding on approach, and finding qualified translators (ML).
2	Forward translation	Performing two independent forward translations (JB1, ML).
3	Reconciliation	Reconciliation of two forward translations into one consensus translation (ML, JB1).
4	Backward translation	Performing a backward translation of INCRESE-NL to English using a Dutch native speaker with proficient English skills, who was blinded to the original INCRESE (MR).
5	Revision of reconciled forward translation	Assessing semantic equivalence of backward translation with the original INCRESE, identification of issues in the reconciled translation, tracking of changes, and implementation in a revised version (ML, JB1).
6	Proofreading	Four proof readers, which included people with lived experience (SC, IvB, JB2, MR) checking the translation on Dutch language errors, readability of sentences, on overall impression on the feasibility and acceptability of the instrument, and on anything missing. Creating the draft version of INCRESE-NL to be validated.
7	Instrument testing/ formal evaluation	Evaluation of feasibility, validity, and reliability of INCRESE-NL by testing the instrument with six coders (SC2, NvS, MdV, MR, JT, MS) of different professional backgrounds, all coding 30 recovery narratives after receiving training on the use of INCRESE-NL.

8	Post-testing/formal evaluation review (Analysis/Revisions)	Calculating the intercoder reliability, collecting the evaluations of coders on feasibility and user friendliness of INCRESE-NL. Discussions on results within the research team (JB1, ML, SC, IvB).
9	Documentation	Documentation of experiences of coders while using INCRESE-NL including advise for future coders (ML; see *Appendix II*).
10	Report	Writing the final summary report, documenting the development of translation, including descriptions of translation and cultural adaptation decisions (ML, JB1, SC, IvB).
11	Archiving/ Record-keeping	Archiving of: qualifications and experience of translation team, documentation of changes made throughout the translation work and rationale for changes, translation report including results of formal evaluation of INCRESE-NL (ML).

Part B: Validation phase of INCRESE-NL

In the validation phase (see Step 7, *Table 1*), INCRESE-NL was tested and evaluated by six coders who each read 30 mental health-oriented (recovery) narratives of the Story Bank Psychiatry[13] and filled in INCRESE-NL for each narrative. The aim was to evaluate the *feasibility*, *validity,* and the *reliability* according to the coders of INCRESE.

Sample size calculation

We ran a sample size calculator[14] including the minimum reliability (ρ0=0.4), expected reliability (ρ1=0.65), significance level (α=0.05), power (80%), six coders and an expected drop-out rate (10%), based on the psychometric results of the original INCRESE. The results suggested a minimum of 28 recovery narratives with six coders, accounting for a 10% dropout rate. We chose to conservatively include 30 recovery narratives.

Recovery narratives

We used 30 text-based (recovery) narratives of the Story Bank Psychiatry (In Dutch: Verhalenbank Psychiatrie),[13] an initiative by the Psychiatry Department of University Medical Centre Utrecht, which is a collection of mental health-oriented narrative experiences from patients, caregivers, and professionals. The website contains written mental health-oriented narratives with a variation in demographics, content, and diagnoses. Out of a database of over 200 narratives, we chose to include summarized versions of narratives. Informed consent was provided for usage of the narratives for scientific research and for anonymous publication on the open available website. The included 30 narratives were written from a patient perspective. As a strategy to maximize variation, we included narratives with variation in gender, age, different self-reported diagnoses, and narratives using a non-diagnostic framework.

Coders

To check for eligibility of use by different potential users, we included clinicians (n=2), researchers (n=2), and lived experience experts (n=2) as coders in the study. The instructions on how to use INCRESE were translated and adapted to the Dutch situation and offered to the coders during a training session. To maintain the coder's wellbeing while coding narratives, a list of strategies for dealing with the potential

negative effects of narratives was shared with the coders. After the training, coders were asked to what extent they felt confident about their ability to use INCRESE-NL on a 5-point Likert scale (1=not confident at all; 5=very confident). Additional training was offered if the coders scored a 3 (doubtful) or lower.

Measures of validation

To examine the *feasibility* of INCRESE-NL according to coders, we evaluated whether the instrument suits the need for sustained and routine use and if it is meaningful for its purpose.[15] In addition, we examined the *user-friendliness* of INCRESE-NL by checking whether the language of the instrument is appropriate in combination with an acceptable administrative, physical, and emotional burden, while preserving the quality of the categorization of the recovery narrative.

Content validity was examined both qualitatively and quantitatively. All coders described whether the items in each section are a good representation to determine these characteristics in the narratives and assigned a rating on a 5-point-Likert scale (1=strongly disagree about good representation; 5=strongly agree about good representation) with ≥4 being a sufficient content validity.

To check for consistency in coding of the INCRESE-NL items, we assessed the *intercoder reliability* between coders.

Data analysis

As we included only text-based narratives from the Story Bank Psychiatry, some items will show high to perfect agreement, but low prevalence in different answer categories. Fleiss Kappa is less suitable for datasets with high agreement and low distribution due to the way it performs chance correction, also referred to as the "Kappa paradox." Fleiss Kappa values drop disproportionately in this situation, even with only one or a few disagreements. Therefore, the Gwet's AC1 agreement coefficient was chosen as our main measure to calculate intercoder reliability. Gwet's AC1 is considered a stable intercoder reliability measure in the above-mentioned situation .[19] For comparison purposes, we did still calculate Fleiss Kappa [20] values (κ) for each individual item. We considered Gwet's AC1 and Kappa values of ≥0.40 to be sufficient.[21] For full transparency of the intercoder reliability scores and effect of chance corrections on the statistical values, we also chose to report the absolute pairwise percent of agreement. Absolute percent agreement does not have clear cut-off values for interpretation; but following guideline indications in the literature,[22, 23] we consider <75% a low percentage of agreement, ≥75-90% as acceptable, and >90% as high agreement. Of note, to randomize for any remaining effects of learning how to use INCRESE-NL or fatigue during the coding process, narratives were offered to the coders in a random order using a random sequence generator for narrative number (Smallest value: 1; Largest value: 30),[24] after which the sequence per coder was recoded to coder ID and numbered 1 to 30. Coders were asked to follow their order of numbering of the narratives during the coding process. An average content validity index score per section was calculated from the scores on a 5-point Likert scale. IBM SPSS Statistics (Version 29) and RStudio (Version 2022.07.2, Build 576) with the irrCAC package[25] were used for quantitative statistical analyses. Written survey evaluations were collected via a Qualtrics survey.[26] Line-by-line coding was performed to find patterns and other relevant feedback regarding the overall impression, feasibility, and user friendliness of the INCRESE-NL. Summaries of the answers are presented in the results section.

Results

Part A: Translation phase

In Step 1 (Preparation), translation permission was obtained, and translators were identified. In Step 2 (Forward translation), independent forward translation developed INCRESE-NL v1A and v1B. In Step 3 (Reconciliation), v1A and v1B were harmonized to produce INCRESE-NL v2. This involved discussing differences and reaching consensus on the best fitting Dutch translation. In Step 4 (Backward translation), a back translation of the INCRESE-NL v2 was performed. In Step 5 (Revision of reconciled forward translation), the back translation was assessed for semantic equivalence with the original instrument and changes were made on a word-level to INCRESE-NL v2. In Step 6 (Proofreading), proofreading on the feasibility, user-friendliness, and cultural relevance led to adaptations in three items. The answer categories of item 13 (Ethnicity) were adapted to the Dutch population and based on migration background statistics. Similarly, item 32 (Positioning) was slightly changed to be more fitting for the organization of services in the Netherlands and item 34 (Relationship to recovery) was slightly altered to ensure the intended meaning of the original item was maintained. A final version of the INCRESE-NL was subsequently created, ready for further validation (*see Appendix I)*.
After the translation phase, the coders were trained in using INCRESE-NL. Four out of the six coders reported feeling (very) confident (n=4) in their ability to use INCRESE-NL. Two coders were less confident (score≤3). After additional inquiry, both coders reported uncertainty in coding specific items that are open to multiple interpretations, for instance, item 31 (Genre) and 33 (Tone). However, they reported feeling confident enough to use the instrument without further training.

Part B: Validation phase of INCRESE-NL

Qualitative summary

Feasibility
All coders completed the evaluation survey after coding 30 recovery narratives. The overall impression regarding the feasibility of INCRESE-NL was that the instrument is suitable to make a relatively complete inventory of the most important characteristics of recovery narratives. For applications in clinical practice, the instrument was considered useful to be able to identify narratives with certain characteristics or content that individual recipients might (not) wish to read (e.g., item 34 ‘Relationship to recovery’ = ‘Making progress’ or ‘Living well’) or avoid (e.g., ‘Content warning’ = ‘Abuse or sexual violence’). Most items were deemed to be clear and user-friendly, but some items were found to be open to multiple interpretations. In these instances, coders were missing nuances in the answer options, especially when it was obligatory to choose one answer (i.e., item 14 ‘Stage of recovery’; item 31 ‘Genre’; Item 32 ‘Positioning’; Item 33 ‘Tone’; Item 34 ‘Relationship to recovery’). One of the coders mentioned that if the goal is to collect only a rough selection of characteristics, the instrument contains too many questions. See *Appendix II* for a list of experiences of the coders using INCRESE-NL, including advice for future coders.

User-friendliness
Coders noted that language use was appropriate (n=3), the instrument form was easy to fill out (n=3), and that the reading and coding process was quite time

consuming (time range: 8:24 – 46:49 minutes per narrative; length of narrative, range: 729 – 1256 words per narrative), but doable (n=3). The average time spent coding per narrative varied per coder from 12:00 – 33:58 minutes. One coder mentioned that coding multiple narratives in a row can lead to mix-ups and advised not to code more than three at a time. Another coder reported that one of the included narratives slightly triggered his/her own trauma.

Content validity evaluation
For the qualitative content validity evaluation, the coders were asked to what extent they found the items and descriptions to be suitable for capturing information from the narratives for the corresponding section topics.

For *Sections 1 (Narrative eligibility) and 2 (Narrative mode)*, the overall opinion was that the items are clearly defined and able to capture the most important elements of a recovery narrative. Although the coders only rated text-based narratives, they expect the INCRESE-NL will also capture information of other modalities.

For *Section 3 (Narrator characteristics)*, most items were considered to be clear and contributed to understanding characteristics of the narrator. One coder mentioned it is difficult to determine when there is enough content in the narrative to give the correct answer for some items, even though it seemed a straightforward item at first sight (such as item 16 'Sexuality'). The original INCRESE instructions did cover what to do in cases like these, so we reiterated the coding instructions for some of these items in *Appendix II*. For the items regarding diagnosis, one of the coders reported to being unsure whether only primary or also secondary diagnoses should be checked off, when the diagnosis hierarchy is identified as such by the narrator. Another point of attention was that the description of a symptom, such as 'hearing voices' was in both the diagnosis item (item 27) and the non-diagnostic framework item (item 30).

Section 4 (Narrative characteristics) was rated as the most difficult to code. Coders mentioned that for some items (i.e., Tone and Trajectory) the answer categories are not able to summarize the complex dynamics of a narrative. They found it difficult to choose one best matching answer for these items. One coder mentioned that the answer categories for item 31 (*Genre)* do not match well with the description, which also makes it more difficult to choose an answer. Another coder mentioned that Section 4 is the most subjective section.

The items of *Section 5 (Content warnings)* were evaluated as being suitable but also as being more ambiguous to code than the items in Sections 1-3. One coder mentioned that there seems to be overlap between items 37 and 40, which are both about types of violence and aggression.

The overall impression of *Section 6 (Turning points)* was that checking off one or more turning points was difficult, as there was some confusion regarding how explicit a turning point needs to be mentioned in the narrative. Also, one of the coders perceived two item descriptions as overlapping (i.e., *Self-acceptance* and *Shift in*

identity) and therefore to be more open for interpretation. The other turning point items descriptions were found easier to distinguish.

Coders reported that the items and descriptions of *Section 7 (Narrative content)* give a complete overview of the content of the narrative. The main reported issue was the difficulty in determining whether a topic was significantly present, meaning the item should be checked, or the topic was just mentioned briefly, in which case the item should not be checked. One coder remarked that he/she missed the topics 'relationship to society,' both socially and in terms of possibilities for participation or reintegration, and 'availability and access to services,' considering the circumstances of complexity and comorbidity of diagnoses in relation to the way the Dutch mental health care system is organized and standardized.
Overall, the current version of INCRESE-NL was deemed suitable to capture the most important characteristics of recovery narratives and was evaluated as user-friendly. A summary of all the experiences that the coders reported in using the instrument (*see Appendix II*) can serve as background information for future users.

Quantitative analysis

Content validity index
The overall mean content validity score for INCRESE-NL was 4.26 (SD=0.33), indicating the instrument is rated to have a sufficient overall content validity. Sections 4 'Narrative characteristics' (M=3.50, SD=0.84) and 5 'Content Warnings' (M=3.83, SD=0.98) had insufficient scores, while Section 1 'Narrative eligibility' (M=4.67, SD=0.52) and Section 2 'Narrative mode' (M=5.00, SD=0.00) had the highest content validity scores.

Intercoder-reliability
For ten out of 77 items, the Gwet's AC1 coefficient and Fleiss Kappa could not be calculated because all ratings were the same for these items. Of the remaining 67 items, 62 items had a sufficient Gwet's AC1 coefficient (AC1 ≥ 0.40), and five item scores were insufficient (AC1 <0.40). For the Fleiss Kappa calculation, 39 of these 67 items had a sufficient Kappa value (κ ≥ 0.40), and 28 items were insufficient (κ < 0.40).

Ten of the 77 instrument items scored 100% on percentage of absolute agreement as all ratings were the same. Of the remaining 67 items, 49 items had acceptable or high percentage of absolute agreement, and 18 items had low percentage of absolute agreement. Seventeen items with insufficient Kappa values had acceptable or high percentage of agreement scores (range: 78% to 99%) and sufficient Gwet's AC1 coefficients (range: 0.70-0.99). The intercoder agreement and reliability coefficients are shown in *Table 2*.

Table 2 *Intercoder-reliability of INCRESE-NL on item-level*

Item		Percentage agreement	Gwet's AC1	Fleiss Kappa current study	Fleiss Kappa original study
SECTION 1: NARRATIVE ELIGIBILITY **[In Dutch: ONDERDEEL 1: GESCHIKTHEID VAN HET NARRATIEF]**					
1.	Lived experience account? *[In Dutch: Weergave persoonlijke ervaringen]*	100%[a]	N/A[a]	N/A[a]	0.79
2.	Is this a narrative? *[In Dutch: Is dit een narratief?]*	98%	0.98	**0,19**	0.85
3.	Narrator-defined adversity *[In Dutch: Bevat het narratief tegenslag of worsteling?]*	100%[a]	N/A[a]	N/A[a]	0.62
4.	Narrator-defined success *[In Dutch: Bevat het narratief succes, krachten en overleving?]*	95%	0,94	**0,38**	0.75
SECTION 2: NARRATIVE MODE **[In Dutch: ONDERDEEL 2: NARRATIEF MODUS]**					
5.	Written elements *[In Dutch: Geschreven elementen]*	100%[a]	N/A[a]	N/A[a]	0.99
6.	Sound elements *[In Dutch: Geluidselementen]*	100%[a]	N/A[a]	N/A[a]	0.99
7.	Moving image elements *[In Dutch: Bewegende beelden]*	100%[a]	N/A[a]	N/A[a]	1.00
8.	Static image elements *[In Dutch: Statische afbeeldingen]*	99%	0,99	**-,01**	1.00
9.	Total words (#) *[In Dutch: Totaal aantal woorden]*	100%[a]	N/A[a]	N/A[b]	N/A
10.	Total length (mins) *[In Dutch: Totale lengte (min)]*	100%[a]	N/A[a]	N/A[a]	N/A
SECTION 3: NARRATOR CHARACTERISTICS **[In Dutch: ONDERDEEL 3: KENMERKEN VAN DE VERTELLER]**					
11.	Gender *[In Dutch: Gender]*	97%	0,97	0,94	0.90
12.	Age *[In Dutch: Leeftijd]*	96%	0,95	0,92	0.89
13.	Ethnicity *[In Dutch: Etniciteit]*	92%	0,92	**0,27**	0.95
14.	Stage of recovery *[In Dutch: Herstelfase]*	**70%**	0,66	**0,33**	0.88[c]
15.	Location of the narrator *[In Dutch: Locatie van de verteller]*	89%	0,87	**0,15**	0.91
16.	Sexuality *[In Dutch: Seksualiteit]*	**69%**	0,63	**0,38**	0.88
17.	Visual difficulties *[In Dutch: Visuele beperking]*	100%[a]	N/A[a]	N/A[a]	1.00
18.	Hearing difficulties *[In Dutch: Gehoorproblemen]*	100%[a]	N/A[a]	N/A[a]	1.00
19.	Mobility/stamina difficulties *[In Dutch: Beperkt(e) mobiliteit, uithoudingsvermogen]*	95%	0,94	0,56	0.71

20.	Cognitive difficulties *[In Dutch: Cognitieve problemen]*	92%	0,89	0,61	0,75[d]
21.	Self-care difficulties *[In Dutch: Problemen bij zelfzorg]*	98%	0,98	0,79	0.76
22.	Neuro-developmental related *[In Dutch: Neuro-ontwikkeling gerelateerd]*	96%	0,93	0,88	0.57
23.	Eating or food-related *[In Dutch: Eet- of voedingsgerelateerd]*	96%	0,96	0,72	0.94
24.	Mood-related *[In Dutch: Stemmingsgerelateerd]*	85%	0,70	0,69	0.87
25.	Personality-related *[In Dutch: Persoonlijkheid gerelateerd]*	92%	0,90	0,41	0.89
26.	Obsessive-compulsive related *[In Dutch: Obsessief-compulsief gerelateerd]*	96%	0,96	0,45	0.75
27.	Schizophrenia or other psychosis-related *[In Dutch: Schizofrenie, psychose gerelateerd]*	83%	0,71	0,59	0.92
28.	Trauma/stress-related *[In Dutch: Trauma, stress gerelateerd]*	94%	0,92	0,73	0.60[d]
29.	Substance-related *[In Dutch: Middelen-gerelateerd]*	89%	0,86	0,54	0.71
30.	Uses a non-diagnostic framework *[In Dutch: Gebruikt een non-diagnostisch kader]*	78%	0,70	**0,10**	0.79[d]
SECTION 4: NARRATIVE CHARACTERISTICS **[In Dutch: ONDERDEEL 4: EIGENSCHAPPEN VAN HET NARRATIEF]**					
31.	Genre *[In Dutch: Genre]*	**51%**	**0,36**	**0,33**	0.82
32.	Positioning *[In Dutch: Positionering]*	**72%**	0,61	0,47	0.85
33.	Tone *[In Dutch: Toon]*	**54%**	0,41	**0,31**	0.78
34.	Relationship with recovery *[In Dutch: Relatie met herstel]*	**48%**	**0,37**	**0,22**	0.58
35.	Trajectory *[In Dutch: Traject]*	**64%**	0,57	**0,38**	0.83
36.	Use of metaphor, symbolic language *[In Dutch: Gebruik van metafoor of symbolisch taalgebruik]*	**55%**	**0,10**	**0,10**	0.82
SECTION 5: CONTENT WARNINGS **[In Dutch: ONDERDEEL 5: INHOUDSWAARSCHUWINGEN]**					
37.	Abuse or sexual violence *[In Dutch: Mishandeling of seksueel geweld]*	87%	0,75	0,72	0.75
38.	Loss of life or endangerment to life *[In Dutch: Verlies van leven of gevaar voor het leven]*	77%	0,59	0,45	0.77[d]
39.	Self-harm including eating disorders *[In Dutch: Zelfbeschadiging inclusief eetstoornissen]*	82%	0,69	0,55	0.83[d]
40.	Violence or aggression *[In Dutch: Geweld of agressie]*	85%	0,79	0,40	0.66[d]
41.	Injustice, prejudice, discrimination	82%	0,76	**0,28**	0.85

	[In Dutch: Onrechtvaardigheid, vooroordelen en discriminatie]				
SECTION 6: TURNING POINTS					
[In Dutch: ONDERDEEL 6: KEERPUNTEN]					
42.	Taking charge *[In Dutch: Regie nemen]*	84%	0,80	**0,24**	0.62
43.	Interventions/support from others *[In Dutch: Interventies, steun van anderen]*	**71%**	0,43	0,43	0.70
44.	Self-acceptance *[In Dutch: Zelfacceptatie]*	83%	0,79	**0,19**	0.68
45.	Spiritual/existential experience *[In Dutch: Spirituele, existentiële ervaring]*	95%	0,95	0,51	0.81[d]
46.	'Rude awakening' *[In Dutch: 'Ruw wakker geschud worden']*	96%	0,95	0,56	0.97
47.	Shift in identity *[In Dutch: Verschuiving in identiteit]*	98%	0.98	**-0,01**	0.63
SECTION 7: NARRATIVE CONTENT					
[In Dutch: ONDERDEEL 7: INHOUD NARRATIEF]					
48.	Pregnancy/birth *[In Dutch: Zwangerschap, geboorte]*	96%	0,96	**0,22**	0.99
49.	Family *[In Dutch: Familie]*	**59%**	0,46	0,42	0.81
50.	Being in care *[In Dutch: In zorg zijn]*	100%[a]	N/A[a]	N/A[a]	1.00
51.	Education *[In Dutch: Opleiding]*	76%	0,72	0,46	0.87
52.	Friendships *[In Dutch: Vriendschappen]*	**67%**	0,59	0,49	0.79
53.	Relationships *[In Dutch: Relaties]*	77%	0,71	0,65	0.77
54.	Housing *[In Dutch: Huisvesting]*	87%	0,86	**0,29**	0.83
55.	Income *[In Dutch: Inkomen]*	94%	0,94	**0,26**	0.84
56.	Work *[In Dutch: Werk]*	**66%**	0,58	0,45	0.78
57.	Criminal justice system *[In Dutch: Strafrechtelijk systeem]*	91%	0,90	0,43	0.70
58.	Diagnosis *[In Dutch: Diagnose]*	**54%**	**0,39**	**0,35**	0.80
59.	Medication *[In Dutch: Medicatie]*	75%	0,68	0,61	0.86
60.	Relationship with mental health professional *[In Dutch: Zorgrelatie met GGZ-medewerker]*	**46%**	**0,28**	**0,25**	0.83
61.	Peer support *[In Dutch: Lotgenotencontact]*	**64%**	0,58	**0,12**	0.73[e]|0.78[e]
62.	Involuntary use of mental health services *[In Dutch: Onvrijwillig gebruik van GGZ zorg]*	87%	0,84	0,64	0.84
63.	Hospitalization *[In Dutch: Opnames]*	**64%**	0,55	0,44	0.78
64.	Psychological services *[In Dutch: Psychologische hulpverlening]*	**61%**	0,51	**0,34**	0.78
65.	Alternative therapies/healing *[In Dutch: Alternatieve therapieën, heling]*	92%	0,92	0,63	0.82
66.	Being in natural environments	94%	0,94	0,50	0.71

	[In Dutch: In een natuurlijke omgeving zijn]				
67.	Animals/pets *[In Dutch: (Huis)dieren]*	99%	0,99	0,85	0.77
68.	Community activities *[In Dutch: Gemeenschapsactiviteiten]*	94%	0.94	**-0,02**	0.63
69.	Hobbies/interests/creative activities *[In Dutch: Hobby's, interesses, creatieve activiteiten]*	88%	0,85	0,59	0.74
70.	Physical activities *[In Dutch: Lichamelijke activiteiten]*	93%	0,93	**0,09**	0.71
71.	Activism *[In Dutch: Activisme]*	94%	0,93	**0,03**	0.61[d]
72.	Spiritual/religious activities *[In Dutch: Spirituele of religieuze activiteiten]*	94%	0,94	0,64	0.89
73.	Stigma *[In Dutch: Stigma]*	**73%**	0,67	0,46	0.79
74.	Caring responsibilities *[In Dutch: Zorgverantwoordelijkheden]*	85%	0,83	**0,28**	0.58
75.	Family experiences of mental health issues *[In Dutch: Ervaringen met psychische gezondheidsproblemen in de familie]*	81%	0,79	0,41	0.87
76.	Diet/nutrition *[In Dutch: Eetpatroon, voeding]*	95%	0,95	0,56	0.81
77.	Volunteering *[In Dutch: Vrijwilligerswerk]*	88%	0,86	0,61	0.89

bold. low percentage of agreement, or below predefined item-level threshold
a. All responses were the same
b. All responses are based on word count in Microsoft Word
c. The ordering of this item has changed in refinements of the original INCRESE
d. The naming of this item has slightly changed in the refinements of the original INCRESE
e. The items formal and informal peer support were merged into one item, peer support

Discussion

In this study, we translated and validated the INCRESE for the use of Dutch narratives (INCRESE-NL). Overall, the coders evaluated INCRESE-NL as sufficiently feasible and user-friendly. The content validity was found good for most sections but insufficient for sections 4 and 5. Sections 1-3, 5, and 6 had high intercoder reliability and were evaluated as having mostly clear content and coding rules. Section 4 had poor reliability scores and was reported as the most difficult section to code with items that are open to multiple interpretations, such as item 31 (Genre), item 33 (Tone), and item 35 (Trajectory). Section 7 had two poorly performing items, item 58 (Diagnosis) and item 60 (Relationship with mental health professional) but were otherwise evaluated as having suitable items, albeit with somewhat ambiguous coding requirements. The Gwet's analysis showed us that the reliability of most items (62/67) is sufficient to good. This is supported by the acceptable to high percentages of absolute agreement among coders on most of these items. We therefore conclude that INCRESE-NL is a useful tool to characterize Dutch mental health recovery narratives.

Interpretation of findings

The quantitative analysis of the Gwet's AC1 coefficient demonstrates similar reliability

of INCRESE-NL compared to the original INCRESE instrument[11] for most of the items. The lower scores on some of the above-mentioned items can be explained by several factors related to clarity in item topic definitions and agreements on the coding instructions.

One issue is that the trainers and coders were missing detailed information about certain item topic definitions or nuances in answer categories to be able to give a diligent answer. Another issue to explain inconsistency among coders is uncertainty regarding how much evidence there needs to be in a narrative to answer an item such as *sexuality.* For example, if a male narrator speaks about 'his wife,' one could interpret this as sufficient information to characterize the narrator as heterosexual. However, it could also be argued that the narrator makes no explicit statements regarding his sexuality, and he could be bisexual or a closeted homosexual, and therefore it would be better to leave the item unanswered. This scenario was covered by the original INCRESE coding instructions, and we therefore reiterated the instructions for these items in *Appendix II*. Agreement prior to using INCRESE-NL on exactly when it is legitimate to make assumptions about an answer may help to improve consistency among coders. A third issue is the type of item. Section 4 contains progressive items about latent content such as item 31 (Genre) and item 33 (Tone) of the narrative, which required more interpretation by the coders and were found more difficult to code than the more straightforward manifest items, which was also mentioned by the coders in the original INCRESE's pilot evaluation. The fourth issue is about lack of clarity on how big a part of the narrative about a content topic should be to be identified as significantly present. For instance, if having an animal or pet is mentioned only briefly in a sentence, it was unclear if it should be ticked or left blank. Finally, there was some uncertainty about whether content items should only be coded when the content is directly applicable to the narrator or also when it applies to other people, such as a relative or a child. Therefore, for some items, it may help to set stricter coding rules and instrument instructions and to communicate them by qualified trainers on when and when not to code something. We make several recommendations to this end in *Appendix II*.

A statistical factor of influence is that the Gwet's AC1 analysis treats the items *Stage of recovery* and *Relationship to recovery* as nominal items,[20] even though these items are ordinal with a hierarchical rank order. The analysis we used did not take the degree of agreement between coders into account, only agreement or disagreement in answer responses. Although this leads to slightly more conservative results, it would not have changed our interpretation of results on these items.

Strengths and limitations

This study is the first to adapt and validate INCRESE in another language. A strength is that we followed quality standards through a step-by-step execution of the updated practices of a consortium translation process for translating patient reported outcome instruments.[12] Furthermore, the power analysis on the number of recovery narratives to score for the Dutch validation strengthens the reliability of our results. Another added value is that this study included coders with different perspectives (clinicians, researchers, and lived experience experts). Feasibility and validity were examined both quantitatively (Gwet's AC1, Fleiss Kappa, percentages of agreement among

coders) and qualitatively (written evaluations). Moreover, a training day was organized to train the coders in using INCRESE-NL.

A limitation of this study is that we only used (recovery) narratives that were text-based. Narratives in different formats, such as audio or video, were not available for this study. Consequently, for some items, an intercoder reliability coefficient could not be calculated and only qualitative evaluation was possible. However, the qualitative evaluation supported that INCRESE-NL would likely be able to also capture information of different modalities. Furthermore, the narratives included in this study were all collected by the "Story Bank Psychiatry"[13] with the same interview protocol and a limited pool of interviewers, all conducted in a similar interview setting. This led to narratives with a more similar structure, content, and length (average of 729-1256 words per narrative) than in the original INCRESE study (different sources, modalities, and length). As an example, a study including 71 personal accounts that were shared in different formal and informal settings, showed that the context of sharing a story greatly affects the '*stories that people tell about their stories.*[27] Moreover, different curator goals such as 'fighting stigma' or 'reframing mental illness' may influence decisions on inclusion of more similar narratives in a collection and editing content of narratives.[28] Due to homogeneity in our dataset, we were not able to fully compare the Fleiss Kappa values in the current and original study.

Conclusions

INCRESE-NL has a sufficient feasibility, validity, and reliability and can help to characterize Dutch mental health-oriented narratives. Using INCRESE-NL to characterize content of such narratives may help people to select relevant narratives that may induce positive impact and minimize harm to a recipient. It is useful to complete training in using INCRESE-NL, to make clear agreements on the coding rules, especially for section 4 about narrative characteristics. Therefore, we advise, in line with the original INCRESE instructions, to practise coding with several narratives and to code with multiple coders, assess concordance, and discuss disagreements and emotional impact. INCRESE-NL may be used as a 'diversity tool' to identify gaps in Dutch narrative collections, as well as for characterization of narratives complementary to qualitative analysis in which new areas and emergent themes of recovery can be explored.

Acknowledgements

We thank Klaas Wardenaar, Roland Reezigt and Maarten Brilman for their statistical guidance and Matthijs Ruiter [MR) and Jopie Berg (JB2) for their contributions in the translation phase. Furthermore, we thank Joke Tonn (JT), Matthijs Ruiter (MR), Mirjam Smid (MS), Maartje de Vries (MdV), Nienke van Sambeek (NvS) and Stijn Crutzen (SC2) for their help in coding narratives in the validation phase of INCRESE-NL.

Mike Slade acknowledges the support of the NIHR Nottingham Biomedical Research Centre. SRE receives salary support from the NIHR under its Programme Grants for Applied Research Programme (Programme Grants for Applied Research, Personal experience as a recovery resource in psychosis: Narrative Experiences Online

(NEON) Programme, RP-PG-0615-20016). FN acknowledges support from a NIHR Advanced Fellowship (NIHR302218) and a University of Nottingham Anne McLaren Fellowship.

All authors declare that they have no conflicts of interest.

Table legend

Table 1: INCRESE to INCRESE-NL: Translation Process Steps
Table 2: Intercoder-reliability of the INCRESE-NL on item-level

Appendices

Appendix I, INCRESE-NL instrument
Appendix II, INCRESE-NL, Table of coders' experiences per section and coding advise

References

1. Leamy M, Bird V, Le Boutillier C, Williams J, Slade M. Conceptual framework for personal recovery in mental health: Systematic review and narrative synthesis. Br J Psychiatry. 2011;199(6):445-52.
2. Wood L, Alsawy S. Recovery in psychosis from a service user perspective: A systematic review and thematic synthesis of current qualitative evidence. Community Ment Health J. 2018;54(6):793-804.
3. McGranahan R, Rennick-Egglestone S, Ramsay A, Llewellyn-Beardsley J, Bradstreet S, Callard F, Priebe S, Slade M. Curation of mental health recovery narrative collections: Systematic review and qualitative synthesis. JMIR Mental Health. 2019;6(10):e14233.
4. Yeo C, Rennick-Egglestone S, Armstrong V, Borg M, Franklin D, Klevan T, Llewellyn-Beardsley J, Newby C, Ng F, Thorpe N, Voronka J, Slade M. Uses and misuses of recorded mental health lived experience narratives in healthcare and community settings: Systematic review. Schizophr Bull. 2021
5. Dawson L, Mullan B, Touyz S, Rhodes P. Are recovery stories helpful for women with eating disorders? A pilot study and commentary on future research. Journal of Eating Disorders. 2018;6(1):21.
6. Roe J, Brown S, Yeo C, Rennick-Egglestone S, Repper J, Ng F, Llewelyn-Beardsley J, Hui A, Cuijpers P, Thornicroft G, Manley D, Pollock K, Slade M. Opportunities, enablers, and barriers to the use of recorded recovery narratives in clinical settings. Front Psychiatry. 2020;11:589731.
7. Rennick-Egglestone S, Ramsay A, McGranahan R, Llewellyn-Beardsley J, Hui A, Pollock K, Repper J, Yeo C, Ng F, Roe J, Gillard S, Thornicroft G, Booth S, Slade M. The impact of mental health recovery narratives on recipients experiencing mental health problems: Qualitative analysis and change model. PLoS One. 2019;14(12):e0226201.
8. Ng F, Charles A, Pollock K, Rennick-Egglestone S, Cuijpers P, Gillard S, van der Krieke L, Bongaardt R, Pomberth S, Repper J, Roe J, Llewellyn-Beardsley J, Yeo C, Hui A, Hare-Duke L, Manley D, Slade M. The mechanisms and processes of connection: Developing a causal chain model capturing impacts of receiving recorded mental health recovery narratives. BMC Psychiatry. 2019;19(1):413-z.
9. Ng F, Newby C, Robinson C, Llewellyn-Beardsley J, Yeo C, Roe J, Rennick-Egglestone S, Smith R, Booth S, Bailey S, Castelein S, Callard F, Arbour S, Slade M. How do recorded mental health recovery narratives create connection and improve hopefulness? Journal of Mental Health. 2022;0(0):1-8.
10. Llewellyn-Beardsley J, Rennick-Egglestone S, Callard F, Crawford P, Farkas M, Hui A, Manley D, McGranahan R, Pollock K, Ramsay A, Sælør KT, Wright N, Slade M. Characteristics of mental health recovery narratives: Systematic review and narrative synthesis. PLoS One. 2019;14(3):e0214678.
11. Llewellyn-Beardsley J, Barbic S, Rennic-Egglestone S, Ng F, Roe J, Hui A, Franklin D, Deakin E, Hare-Duke L, Slade M. INCRESE: Development of an inventory to characterize recorded mental health recovery narratives. Journal of Recovery in Mental Health. 2020;3(2):25-44.
12. Eremenco S, Pease S, Mann S, Berry P. Patient-reported outcome (PRO) consortium translation process: Consensus development of updated best practices. J Patient Rep Outcomes. 2017;2(1):12.

13. Verhalenbank Psychiatrie. cited Oct 28, 2021]. Available from: https://psychiatrieverhalenbank.nl/
14. Arifin WN. wnarifin.github.io > Sample size calculator. cited Mar 22, 2022]. Available from: https://wnarifin.github.io/ssc/ssicc.html
15. M S, G T, G G. The feasibility of routine outcome measures in mental health. Social psychiatry and psychiatric epidemiology. 1999;34(5)
16. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of cohen's kappa and gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. BMC Medical Research Methodology. 2013;13(1):61.
17. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. the problems of two paradoxes. Journal of Clinical Epidemiology. 1990;43(6):543-9.
18. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. resolving the paradoxes. Journal of Clinical Epidemiology. 1990;43(6):551-8.
19. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. British Journal of Mathematical and Statistical Psychology. 2008;61(1):29-48.
20. Fleiss JL. Measuring nominal scale agreement among many raters. Psychological Bulletin. 1971;76:378-82.
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.
22. Hartmann DP. Considerations in the choice of interobserver reliability estimates. J Appl Behav Anal. 1977;10(1):103-16.
23. Stemler SE. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. Practical Assessment, Research, and Evaluation. 2004;9(4)
24. Haahr M. Random Sequence Generator. cited Jan 11, 2023]. Available from: https://www.random.org/sequences/
25. Gwet KL. irrCAC: Computing Chance-Corrected Agreement Coefficients (CAC). Comprehensive R Archive Network (CRAN) cited Nov 8, 2022]. Available from: https://CRAN.R-project.org/package=irrCAC
26. Linda Rich. Qualtrics - Survey Tool. ; 2022. Available from: https://libguides.bgsu.edu/qualtrics.
27. Llewellyn-Beardsley J, Rennick-Egglestone S, Pollock K, Ali Y, Watson E, Franklin D, Yeo C, Ng F, McGranahan R, Slade M, Edgley A. 'Maybe I shouldn't talk': The role of power in the telling of mental health recovery stories. Qual Health Res. 2022;32(12):1828-42.
28. Yeo C, Rennick-Egglestone S, Armstrong V, Borg M, Charles A, Duke LH, Llewellyn-Beardsley J, Ng F, Pollock K, Pomberth S, Walcott R, Slade M. The influence of curator goals on collections of lived experience narratives: A qualitative study. J Recovery Ment Health. 2021; 4(2):16-28.