Journal of Occupational Rehabilitation (2019) 29:325–335
https://doi.org/10.1007/s10926-018-9790-x

Cross-Cultural Adaptation, Reliability and Validity of the Danish Version of the Readiness for Return to Work Instrument

Christina Malmose Stapelfeldt[1,2] · Anne-Mette Hedeager Momsen[1,2] · Thomas Lund[3,4] · Therese Koops Grønborg[5] · Sheilah Hogg-Johnson[6,7,8] · Chris Jensen[9,10] · Janne Skakon[11] · Merete Labriola[1,2]

Published online: 6 June 2018

Abstract

The objective of the present study was to translate and validate the Canadian Readiness for Return To Work instrument (RRTW-CA) into a Danish version (RRTWDK) by testing its test–retest and internal consistency reliability and its structural and construct validity. Cross-cultural adaptation of the six-staged RRTW-CA instrument was performed in a standardised, systematic five-step-procedure; forward translation, panel synthesis of the translation, back translation, consolidation and revision by researchers, and finally pre-testing. This RRTW-DK beta-version was tested for its psychometric properties by intra-class correlation coefficient and standard error of measurement (n = 114), Cronbach's alpha (n = 471), confirmatory factor analyses (n = 373), and Spearman's rank correlation coefficient (n = 436) in sickness beneficiaries from a municipal employment agency and hospital wards. The original RRTW-CA stage structure could not be confirmed in the RRTWDK. The psychometric properties were thus inconclusive. The RRTW-DK cannot be recommended for use in the current version as the RRTW construct is questionable. The RRTW construct needs further exploration, preferably in a population that is homogeneous with regard to cause of sickness, disability duration and age.

Keywords Rehabilitation · Vocational · Reproducibility of results · Sick leave · Surveys and questionnaires · Validation studies

✉ Christina Malmose Stapelfeldt
chrstp@rm.dk

Anne-Mette Hedeager Momsen
anne-mette.momsen@stab.rm.dk

Thomas Lund
thomas.lund.01@regionh.dk

Therese Koops Grønborg
therese@ph.au.dk

Sheilah Hogg-Johnson
shoggjohnson@iwh.on.ca

Chris Jensen
chris.jensen@air.no

Janne Skakon
Janne.skakon@psy.ku.dk

Merete Labriola
merete.labriola@rm.dk

[1] Section for Clinical Social Medicine and Rehabilitation, Department of Public Health, Aarhus University, P.P. Oerums Gade 11, Building 1B, 8000 Aarhus C, Denmark

[2] DEFACTUM, Social and Health Services and Labour Market, Central Region Denmark, MarselisborgCentret, P.P. Oerums Gade 11, Building 1B, 8000 Aarhus C, Denmark

[3] Centre for Social Medicine, Frederiksberg and Bispebjerg Hospital, Nordre Fasanvej 57, 2000 Frederiksberg, Denmark

[4] Department of Public Health, Aarhus University, Bartholins Allé 2, 8000 Arhus C, Denmark

[5] Section for Biostatistics, Department of Public Health, Aarhus University, Bartholins Allé 2, 8000 Aarhus C, Denmark

[6] Institute for Work and Health, 481 University Avenue, Suite 800, Toronto, ON M5G 2E9, Canada

[7] Canadian Memorial Chiropractic College, Toronto, Canada

[8] Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

[9] Department of Public Health and Nursing, Norwegian University of Science and Technology, Håkon Jarls gate 11, Øya Helsehus, Trondheim, Norway

[10] National Advisory Unit on Occupational Rehabilitation, Haddlandsvegen 20, 3864 Rauland, Telemark, Norway

[11] Department of Psychology, Work and Organisational Psychology, University of Copenhagen, Øster Farimagsgade 2A, 1353 Copenhagen K, Denmark

Introduction

Sick leave from work is often unavoidable, but it is wasteful and damaging to individuals and their families, employers and to the wider society when unduly prolonged [1]. Return to work (RTW) interventions have therefore attracted scientific and political attention [1, 2]. Multiple RTW interventions in individuals on sick leave due to musculoskeletal [3] or mental disorders [4, 5] as well as cancer [6] have revealed inconsistent results regarding efficiency and content. This inconsistency may be explained partly by a poor fit between the intervention content and the needs of the individuals on sick leave targeted by the intervention [7]. Furthermore, research has stressed that the impact of risk factors may vary across different phases of the disability and RTW process [8–10]. The Stages of Change Model [11–16], which focuses on motivation for behavioural change, has been used to describe motivational and behavioural progress needed to resume work [8, 9]. Using this model, Franche et al. proposed that work resumption could be understood as a multi-stage process [8, 17]. This understanding reflects the theoretical framework of the Readiness for RTW (RRTW-CA) instrument developed by Franche and Krause [8]. This instrument claims to cover a number of existing theories of behaviour; i.e. self-efficacy [18] and decisional balance [15] as well as the Phase Model of Occupational Disability, which emphasizes the temporal and developmental aspects of the recovery process [19].

These individual psychological dimensions reflect the worker's overall readiness to resume work. This readiness is influenced also by the interaction of the sick-listed employee with the workplace, the healthcare system and the insurance system, which emphasizes the interpersonal dimension of the RTW process [8]. The RRTW-CA instrument addresses the motivational factors contributing to RTW behaviour and maintenance of work participation [8]. The rationale behind the RRTW-CA instrument is that RTW is a conscious behaviour. Accordingly, a person progresses through stages of change, shifting from the intention not to engage in RTW behaviour in a foreseeable future to the intention and ability to RTW in a sustainable way [8, 17].

To be able to analyse the demand for efficient and individually tailored RTW interventions, the RRTW-CA instrument was developed to help professionals in occupational rehabilitation design efficient RTW interventions. Using the Canadian instrument in a Danish context required translation as well as cross-cultural and conceptual adaptation [20], because the performance of a questionnaire may differ between populations and cultures [20, 21]. To the best of our knowledge, the RRTW-CA instrument has been translated and validated only into the 16-item Norwegian version (RRTW-NO) [22]. In the Norwegian study, Braathen et al. used exploratory factor analysis, thereby identifying a different number of stages with fewer items than used in the Canadian instrument [22]. Yet another Canadian validation study has been conducted based on the original RRTW-CA instrument [23]. In this study, Park et al. were unable to confirm the original structure proposed by Franche et al.; and like Braathen et al., they proposed yet another scoring model with fewer stages and items.

Therefore, the present study has a fivefold aim:

First; to translate and cross-culturally adapt the Canadian instrument (RRTW-CA) into a Danish version (RRTW-DK).

Second; to evaluate the test–retest reliability of the RRTW-DK instrument.

Third; to analyse whether the internal consistency reliability of the RRTW-DK instrument would be better if it was based on the more recent Norwegian scoring model (two + two stages) than the Canadian scoring model (four + two stages). It was hypothesized that the Norwegian scoring model would show superior internal consistency reliability in a Danish setting due to comparable social security schemes and cultures in general.

Fourth; to investigate the structural validity of the RRTW-DK instrument using both the Canadian and the Norwegian scoring models.

And fifth; to explore the construct validity by testing the hypotheses of associations between all stages of the RRTW-DK instrument and the mean score of the Danish 19-item return-to-work self-efficacy instrument (RTWSE-19) [24].

Materials and Methods

Description, Structure and Scoring of the RRTW Instrument

The 22-item RRTW-CA instrument has four underlying stages for persons not working (Part A, 13 items) and two stages for persons currently working part-time or full-time (Part B, 9 items). Items are scored on a five-point ordinal scale (1 = strongly disagree, 5 = strongly agree), with responses to item A7 and B8 reversed for scoring ("Appendix"). Mean scores are calculated for each stage. Stages containing more than 20% missing values are excluded from analysis [21].

Franche and colleagues suggest two ways of operationalising the RRTW-CA instrument; *stage allocation* where participants are allocated to the stage where they have the highest mean score, or a *multidimensional approach* where the mean scores achieved from each stage are maintained and used [17]. In the present study, we chose the

multidimensional approach to keep as much information in test–retest analyses as possible.

The psychometric properties of the Canadian scoring model have previously been analysed and found satisfactory with regards to internal validity and concurrent validity [17]. However, in the Norwegian cross-cultural validation of the Canadian scoring model, the four + two stage structure could not be confirmed [22]. Performing exploratory factor analyses, identified a different number of stages which were assessed by fewer items than used in the Canadian scoring model. Braathen et al. suggested two stages for those not yet returned to work: *RTW inability* (item A1, A2, A4, A5 and A13) and *RTW uncertainty* (item A10, A11 and A12); and two stages for those returned to work: *Uncertain work maintenance* (item B1, B3, B5, B6 and B7) and *Proactive work maintenance* (item B2, B4 and B9) ("Appendix"). In the Norwegian scoring model, item A4, A5, A10 and B9 are reversed to obtain the highest internal consistency reliability [22].

The Translation and the Cross-Cultural Adaptation Process

The methodology used to cross-cultural adapt the Canadian instrument to Danish followed a systematic five-step procedure introduced by Beaton et al. [25]; i.e. forward translation (step I), panel synthesis of the translation (step II), back translation (step III), consolidation and revision by an expert committee of occupational rehabilitation researchers (step IV), and finally pre-testing (step V). Throughout the cross-cultural adaptation process, face validity was evaluated by an expert committee consisting of occupational rehabilitation researchers from DEFACTUM; Social and Health Services and Labour Market, Central Region Denmark; National Advisory Unit on Occupational Rehabilitation, Rauland, Norway; and social workers from Silkeborg Municipality, Denmark. Qualitative analyses of the comments provided by the participants in the pre-test were also performed by the expert committee.

The Canadian instrument was translated into Danish by bilingual translators with permission from Renée-Louise Franche, one of the developers of the Canadian instrument.

The pre-test (step V) was performed to evaluate comprehensibility, usability and content validity of the RRTW-DK instrument. Pre-test participants were identified in two settings. Pre-test of the four stages of the RRTW-DK instrument was carried out at a municipal employment agency among 40 individuals sick-listed for a minimum of 8 weeks. Participants were 18–65 years of age and were invited to participate by a social worker at a sickness case management meeting. Pre-test of the remaining two stages of the RRTW-DK instrument was carried out among 38 individuals, 18–65 years of age. They had been working (part-time or full-time) for approximately 4 weeks following a sick leave spell. The Human Resource Unit of the Region of Southern Denmark invited these participants, who all were public health employees.

Immediately after completing the RRTW-DK instrument, pre-test participants responded to questions about the layout, wording of instructions, wording of items, whether any missing aspects were identified and their overall impression of the instrument. The participants' written comments were used as full text data to evaluate comprehensibility, usability and content validity of the RRTW-DK instrument. Any items that were ambiguous or gave rise to misunderstandings were discussed by the expert committee leading to decisions about whether changes in the questionnaire were necessary. Finally, a synthesis report was written on the identified issues and how they were dealt with.

Following completion of the five-step cross-cultural adaptation process, a RRTW-DK beta-version was evaluated regarding reliability and validity to determine whether there was evidence to support comparable and adequate measurement properties [25–28].

Participants

The RRTW-DK instrument was completed by 374 sickness beneficiaries from a municipal employment agency and 97 out-patients from three hospital wards in the Central Region Denmark. In Denmark, the municipality is responsible for managing cases of sickness absence lasting more than 8 weeks; sick-listed (full-time and part-time) citizens are by law obliged to attend follow-up meetings at the municipal employment agency. At the first meeting, which is scheduled after eight sick leave weeks, the baseline questionnaire used in the present study was administered (T0). At the hospital, patients were invited to participate only if they had been sick-listed (full-time and part-time) for at least 8 weeks. The part-time sick-listed participants responded to part B as they were working. In both settings, inclusion criteria were age above 18 years and ability to speak and understand Danish. Participants recruited from the municipal employment agency also had to be currently employed.

Participants were invited by a social worker at the municipal employment agency and a clinical assistant in the hospital wards. Details of the study were described to the participants, and the inviting person addressed any questions and concerns they might have. Participation was voluntary.

The municipal employment agency participants filled out the questionnaire immediately after having given oral consent (T0, baseline). Participants were asked to complete the questionnaire again at T1 (7 days after T0) using an online questionnaire. A reminder was sent if participants had not responded within 4 days of T1.

The hospital participants were handed both T0 and T1 questionnaires whenever they had an appointment at the hospital. The T0 questionnaire was completed at the hospital, and the T1 questionnaire was completed at home and returned to the research team in a pre-paid envelope. For this, we accepted a maximum of 15 days following T0. Evaluating test–retest reliability, we chose a range of 7–15 days between administrations to reduce the risk of recall of answers and to limit the possibility of changes in the attributes being measured by the RRTW-DK instrument [20].

Additional Questionnaire-Obtained Data

At T0, participants provided information about age, gender, level of post-secondary educational attainment (low < 3 years, medium 3–4 years, high > 4 years), type of work (manual, non-manual or mixed) and whether they suffered from a chronic condition (yes/no). Also, participants filled out the Danish RTWSE-19 questionnaire [24] by responding on a 1–10-point scale if they could overcome a number of RTW barriers (1 = not at all certain, 10 = completely certain). The total mean score was calculated; the higher the score, the higher the self-efficacy. If more than three of the 19 items were missing, the respondent was excluded from analysis.

Statistical Analysis

The study sample was recruited from two different settings (a municipal employment agency and three hospital wards); the study sample was analysed both pooled as one sample and as two separate samples; descriptive statistics for baseline data were produced and tested by Chi square and Wilcoxon rank sum tests.

The number of participants reaching the minimum and maximum mean sum score in stages of readiness was identified; if more than 15% reached these values, floor and/or ceiling effects may be a risk [29]. Floor and ceiling effects were evaluated for the total sample and stratified by recruitment setting.

Test–Retest and Internal Consistency Reliability

Reliability is the degree to which an instrument is free of measurement error [26, 27].

Intra-class correlation coefficients (ICC) [30] and standard error of measurements (SEM) were estimated for the whole sample and stratified on recruitment setting. ICC of at least 0.7 is acceptable, but 0.9 would indicate excellent reliability suggesting that the instrument could be used for individual persons [31].

The internal consistency reliability of the RRTW-DK instrument was assessed using Cronbach's alpha for each stage of readiness. Alpha values between 0.70 and 0.95 are considered sufficient [20]. Cronbach's alpha was evaluated for the total sample and stratified by recruitment setting.

It was hypothesized that the Norwegian scoring model would match the Danish setting better than the Canadian scoring model.

Validity

The dimensionality of the RRTW-DK instrument was evaluated by confirmatory factor analyses (CFA), where items were analysed as categorical measures with a variance-adjusted weighted least-squares method (WLSMV) estimator. Both the Canadian scoring model and the Norwegian scoring model were investigated. Fits of the models were evaluated by fit statistics. Comparative Fit Index (CFI) assesses fit relative to a null model and ranges from 0 to 1, with values between 0.9 and 0.95 indicating acceptable fit and values above 0.95 good fit [32]. Tucker-Lewis Index (TLI) adjusts for the number of model parameters and is interpreted as CFI. Root Mean Square Error of Approximation (RMSEA) expresses the lack of fit per degree of freedom in the model, where values below 0.08 are considered good. Standardised Root Mean Square Residual (SRMR) is the average of the differences between the observed and predicted correlations and ranges from 0 to 1 with values below 0.08 indicating good fit [32].

Construct validity was appraised by the strength of the hypothesized correlation using Spearman's rank correlation coefficient (rho). An absolute value of rho of 0.50 or larger was considered a strong correlation [33]. RTWSE-19 was expected to be negatively correlated with pre-contemplation among those in the non-working sample (Part A) and with uncertain maintenance among those in the working sample (Part B). Participants in these stages do not consider returning to work in the foreseeable future (Part A) or they have reservations toward being able to stay at work (Part B). Thus their self-efficacy regarding work is likely low [8]. Positive correlations were expected for the remaining four stages in part A and B that reflect positive expectancies and actions toward returning to work and staying at work in a sustainable way.

A significance level of $p < 0.05$ was used for all tests. Confirmatory factor analyses were conducted in R version 3.4.1 using the package latent variable analysis (lavaan). All other analyses were performed using the software package Stata/SE 14.2.

Results

The Translation and Cross-Cultural Adaption Process

The aim was to maintain the meaning of the original items. However, some changes were inevitable to improve clarity of

meaning in the Danish language and to adapt the instrument to the Danish culture and setting. The items of the RRTW-CA instrument are shown in "Appendix"; idiomatically challenging items are marked with "*".

In the RRTW-CA instrument, the wording of several questions was either "go back to work", "get back to work", or "return to work", whereas in the forward translation (step I) only two wordings were used: "return to work" and "get back to work".

The panel of occupational rehabilitation researchers discussed several items before agreeing upon a synthesis of the wording for back translation (step III). Consensus was reached to replace the words: "Injury" (B3, B6, B7) and "pain" (B2) with more commonly used Danish words: "health problems" and "discomfort", respectively. Item A5: ("…to build up your strength..") was discussed and changes were made to make the item idiomatically more acceptable in Danish.

A total of 78 participants were included in the pre-test of part A (n = 40) and B (n = 38) (step V). They largely had a positive impression of the comprehensibility, completeness and usability of the test. In their evaluation of the test and the lay-out of the questionnaire, half stated that the questionnaire was "nice, manageable, nice with colours or okay", whereas five commented that it was "messy or confusing".

A majority of participants stated that the instructions were comprehensible. Some would have liked a recall period of a certain duration to consider the questions. The instructions were clarified according to remarks received and the sentence "Please tick the answer (x) that best describes you" was added. Requests regarding a recall period could not be met as this would depart from the instructions in the original version. Furthermore, the instructions were supplemented with "return to work could refer to return to the same job, return to a new job, return to a job at reduced hours or return to the same job but with different responsibilities", since these were common scenarios in the Danish setting. No comments were made concerning the response categories. Though most participants stated that they had no major difficulties in understanding most of the items, some experienced difficulties in understanding the wording of some items.

Some participants mentioned that items A1 and A2 were much alike; several reacted to the negatively formulated questions; and some remarked that item A2 did not make sense. As a consequence A2 was shortened, "As far as you´re concerned" was deleted from the item. The beginning of item A11 "You wish" was considered odd among the pre-test respondents, and the expert committee evaluated these responses as idiomatic discrepancies between Canadian and Danish language and changed the wording to "You lack ideas". No alterations were made in part B.

In general, the participants thought that the questionnaire was usable because the items were relevant and provided insight. More specifically, 17 of the responders to part A and 21 of the responders to part B confirmed that the questionnaire would be useful as a tool for their own preparation and for a conversation with their employer. Five participants stated that the questionnaire was of no use in the RTW process.

Sample Characteristics for Evaluation of Reliability and Validity

A total of 782 participants were invited to participate in the validation study of the RRTW-DK instrument; 471 (60.2%) completed the questionnaire at baseline (T0) with 120 (25.5%) of those responding to the re-test at T1 at a mean of 8.6 days (SD 2.3) from T0.

Non-response analyses revealed no statistically significant differences in the distribution of baseline characteristics between those responding and those not responding at T1, except that fewer responders (n = 120) than non-responders (n = 351) suffered from a chronic disease (results not shown).

The majority of responders were recruited at the municipal employment agency (n = 374); the remaining responders were recruited at hospital wards (n = 97). Significant differences were found between the two samples. Hospital recruits had a significantly higher prevalence of chronic pain and a significantly lower score for stages *prepared for action self-evaluative* and *behavioural* than municipal recruits (Table 1).

Test–Retest and Internal Consistency Reliability

None of the stages reached the ICC threshold of 0.9 for test–retest reliability when analysed for the entire sample (Table 2), although two of the stages surpassed the 0.70 threshold of acceptable ICC—*pre-contemplation* and *uncertain maintenance*. Stratified analyses revealed higher ICCs in general for the job agency recruits than for the hospital recruits, with *pre-contemplation* and *prepared for action self-evaluative* surpassing acceptable threshold of 0.70. The SEMs were between 0.29 and 0.67 with a tendency towards lower SEMs in the population recruited at the job agency than in the hospital population.

The internal consistency reliability of the RRTW-DK instrument, which was evaluated by Cronbach's alpha (α), ranged between 0.44 and 0.81 (Table 3) using the Canadian scoring model. The stages *contemplation* (α = 0.44) and *prepared for action – behavioural* (α = 0.54) were below the threshold of 0.7. In the initial analyses of the internal consistency reliability, *contemplation* had a very low α of 0.32 (not shown in Table 2). However, the translated wording of item A11 was the opposite of what was intended. Reversing the scoring compared to the Canadian scoring so that a low/high scoring still reflected a low/high degree of

Table 1 Baseline characteristics of respondents recruited from a municipal employment agency (n = 374) or hospital wards (n = 97)

		Employment agency		Hospitals	p value
Gender, n (%)					0.65 a
Female		234 (63)		26 (59)	
Male		140 (37)		18 (41)	
Missing		0		53	
Age in years, median (iqr)		45.5 (37–54)		48 (40–53)	0.63 b
Missing, n (%)		0		58	
Education level, n (%)					0.58 a
Low		33 (33)		6 (25)	
Middle		46 (46)		14 (58)	
High		20 (20)		4 (17)	
Missing, n		275		73	
Work type, n (%)					0.56 a
Manual		41 (41)		13 (52)	
Non-manual		34 (34)		6 (24)	
Mixed		24 (24)		6 (24)	
Missing, n		275		72	
Chronic condition, n (%)					**0.01** a
Yes		32 (33)		14 (61)	
No		66 (67)		9 (39)	
Missing, n		276		74	
RTWSE-19, median (iqr)		6.3 (4–8)		6.2 (4–8)	0.40 b
Missing, n		23		11	
RRTW-DK, median (iqr)	n		n		
Pre-contemplation	308	1.0 (1.0–2.0)	73	1.3 (1.0–2.3)	0.05 b
Contemplation		3.7 (3.3–4.0)		3.7 (3.3–4.3)	0.86 b
Prepared for action—self-evaluative		3.0 (2.5–3.3)		2.5 (2.3–3.0)	< **0.001** b
Prepared for action—behavioural		4.0 (3.3–4.3)		3.7 (3.0–4.3)	< **0.01** b
Uncertain maintenance	66	3.2 (2.8–3.6)	24	3.7 (2.9–3.8)	0.27 b
Proactive maintenance		4.3 (4.0–4.5)		4.3 (3.6–4.4)	0.31 b

19-item return-to-work self-efficacy (RTWSE-19)

iqr interquartile range

[a]Chi^2-test, bold value indicates significant ($p < 0.05$)

[b]Wilcoxon rank sum test, bold values indicate significant ($p < 0.05$)

contemplation produced a slight increase in internal consistency reliability for that scale (α = 0.44).

Cronbach's alpha and floor and ceiling effects did not change substantially when stratified on recruitment setting.

When data were analysed using the Norwegian scoring model, Cronbach's alpha ranged from 0.49 to 0.71 (Table 3); thus, only *uncertain maintenance* (α = 0.71) had satisfactory internal consistency reliability.

No floor and ceiling effects were present, except for *pre-contemplation*; approximately 51% had the lowest possible score (Table 3).

Validity

The working sample was too small for CFA and results will be presented only for the non-working sample.

In the Canadian scoring model for non-working, the CFA showed high factor loadings for the stage *pre-contemplation* (range 0.75–0.89). The factor loadings for the three other stages included items with both high and low factor loadings (range 0.40–0.88) with items A12 (from *contemplation*) and A7 (from *prepared for action—self-evaluative*) having negative factor loadings (−0.51 and −0.65); (results not shown). The indices for model fit (Table 4) demonstrated acceptable fit based on two indices (CFI and TLI) and poor fit for the two other indices (RMSEA and SRMR).

The CFA for the Norwegian scoring model for non-working showed both high and low factor loadings (range 0.57–0.83) and a negative factor loading for item A11 (−0.84) from the *uncertainty* stage. The indices for model fit demonstrated good fit based on CFI but poor fit for the remaining indices.

Table 2 Test–retest reliability for the RRTW-DK instrument for individuals currently not working (n = 84) and currently working (n = 30) and stratified on recruitment setting

RRTW-DK	n	T0 median (iqr)	T1	T1–T0	ICC (95% CI)	SEM (95% CI)
Total sample (n = 114)						
Pre-contemplation	84	1.3 (1.0–2.0)	1.2 (1.0–2.0)	0.0 (−0.2 to 0.7)	0.77 (0.68–0.86)	0.38 (0.33–0.44)
Contemplation		3.7 (3.3–4.0)	3.7 (3.3–4.0)	0.0 (−0.3 to 0.3)	0.37 (0.19–0.56)	0.45 (0.34–0.53)
Prepared for action self-evaluative		2.8 (2.4–3.1)	2.8 (2.5–3.3)	0.0 (−0.3 to 0.3)	0.47 (0.30–0.64)	0.42 (0.37–0.50)
Prepared for action behavioural		3.7 (3.0–4.3)	4.0 (3.3–4.3)	0.0 (−0.3 to 0.7)	0.47 (0.31–0.64)	0.55 (0.48–0.65)
Uncertain maintenance	30	3.6 (2.8–3.8)	3.4 (3.0–4.0)	0.0 (−0.2 to 0.2)	0.71 (0.54–0.89)	0.45 (0.36–0.61)
Proactive maintenance		4.3 (3.8–4.8)	4.3 (3.8–4.8)	0.0 (−0.3 to 0.3)	0.44 (0.15–0.73)	0.54 (0.43–0.73)
Job agency-recruited (52)						
Pre-contemplation	40	1.0 (1.0–1.8)	1.0 (1.0–1.7)	0.0 (0.0–0.0)	0.78 (0.65–0.90)	0.38 (0.31–0.49)
Contemplation		3.7 (3.3–4.0)	3.3 (3.3–4.0)	0.0 (−0.3 to 0.3)	0.35 (0.07–0.62)	0.46 (0.38–0.60)
Prepared for action self-evaluative		3.0 (2.5–3.3)	3.0 (2.8–3.4)	0.0 (−0.3 to 0.3)	0.71 (0.56–0.87)	0.29 (0.23–0.37)
Prepared for action behavioural		4.0 (3.3–4.3)	4.0 (3.7–4.3)	0.0 (−0.3 to 0.3)	0.56 (0.34–0.77)	0.45 (0.37–0.58)
Uncertain maintenance	12	3.6 (2.9–3.8)	3.5 (2.4–3.9)	0.0.(−0.3 to 0.0)	0.82 (0.64–1.01)	0.44 (0.31–0.74)
Proactive maintenance		4.5 (4.4–4.9)	4.8 (4.3–5.0)	0.0 (−0.3 to 0.5)	0.68 (0.37–0.99)	0.30 (0.22–0.52)
Hospital-recruited (62)						
Pre-contemplation	44	1.3 (1.0–2.3)	1.3 (1.0–2.0)	0.0 (−0.3 to 0.2)	0.76 (0.64–0.89)	0.37 (0.31–0.47)
Contemplation		3.7 (3.3–4.0)	3.7 (3.3–4.0)	0.0 (−0.3 to 0.3)	0.40 (0.15–0.65)	0.43 (0.36–0.55)
Prepared for action self-evaluative		2.5 (2.3–3.0)	2.7 (2.3–3.0)	0.0 (−0.3 to 0.6)	0.21 (0.00–0.50)	0.52 (0.43–0.66)
Prepared for action behavioural		3.7 (3.0–4.3)	4.0 (3.3–4.3)	0.3 (−0.2 to 0.7)	0.41 (0.16–0.66)	0.63 (0.52–0.80)
Uncertain maintenance	18	3.6 (2.8–3.8)	3.4 (3.0–4.0)	0.1 (0.0–0.3)	0.59 (0.29–0.90)	0.46 (0.35–0.69)
Proactive maintenance		4.1 (3.5–4.3)	4.3 (3.5–4.5)	0.0 (−0.5 to 0.3)	0.27 (0.00–0.70)	0.67 (0.50–1.00)

iqr inter-quartile range, *ICC* intraclass correlation coefficients, *SEM* standard error of measurement

The hypotheses about construct validity were tested using Spearman's rho (Table 5). *Pre-contemplation* and *uncertain maintenance* were the only two stages with ICC values with satisfactory test–retest reliability. As hypothesized, they were both negatively correlated with the global mean score of RTWSE-19. *Uncertain maintenance* was the only stage that reached the threshold of a strong correlation (rho = −0.50, p < 0.0001). The ICCs for the job agency recruits showed a better test–retest reliability than for the hospital recruits, in particular for the stage *Prepared for action—self-evaluative*; this scale had a strong correlation with RTWSE-19 in the total sample (rho = 0.50, p < 0.0001) as well as in the sample of job agency recruits (rho = 0.53, p < 0.0001, Table 5).

The Danish consensus version of the RRTW-DK instrument is available [34].

Discussion

Neither the original Canadian scoring model nor the Norwegian scoring model was confirmed in factor analyses. Test–retest reliability was inconclusive, with only *Pre-contemplation* and *Uncertain maintenance* demonstrating acceptable ICCs. SEM revealed considerable noise between test and retest. Accordingly, the construct validity of the RRTW was questionable.

Test–Retest and Internal Consistency Reliability

The discouraging and inconclusive reliability results invite at least three possible explanations. First, the RRTW-DK instrument could have flaws, i.e. an unstable factor structure could lead to poor reliability. We may be inclined to view this as the explanation, even though the stages of change theory is widely acknowledged [35]. Second, changes unrelated to the instrument may have taken place in the construct between administrations; and third, subject variability in the studied sample may not resemble subject variability in the population the tool is aimed for.

An alternative way to consider variation within responders is to use the stage stability in the stage allocation approach [17], 72% of the responders would be allocated to the same stage at time T0 and T1 (results not shown). Neither Franche et al. [17] nor Braathen et al. [22] investigated what an acceptable, minimal agreement between test and retest should be.

Table 3 Cronbach's alpha, floor and ceiling effects of the RRTW-DK instrument for individuals currently not working (n = 381) and currently working (n = 90)

	Descriptive statistics, baseline					
	n	Cronbach's alpha	Mean (SD)	Median (iqr)	% at floor (95% CI)	% at ceiling (95% CI)
RRTW-DK stages						
Pre-contemplation	381	0.71	1.56 (0.76)	1.0 (1.0–2.0)	50.92 (45.78–56.05)	0.26 (0.01–1.45)
Contemplation		0.44	3.61 (0.56)	3.7 (3.3–4.0)	0.00 (0.00–0.96)	3.41 (1.83–5.76)
Prepared for action—self-evaluative		0.78	2.88 (0.61)	2.8 (2.5–3.3)	0.26 (0.01–1.45)	0.26 (0.01–1.45)
Prepared for action—behavioural		0.54	3.81 (0.80)	4.0 (3.3–4.3)	1.05 (0.29–2.67)	9.19 (6.48–12.54)
Uncertain maintenance	90	0.79	3.29 (0.69)	3.4 (2.8–3.8)	1.11 (0.03–6.04)	0.00 (0.00–4.02)
Proactive maintenance		0.81	4.14 (0.69)	4.3 (4.0–4.5)	2.22 (0.27–7.80)	12.22 (6.26–20.82)
For job agency-recruited only						
Pre-contemplation	308	0.70	1.52 (0.73)	1 (1.0–2.0)	53.25 (47.50–58.92)	0.32 (0.01–1.80)
Contemplation		0.42	3.61 (0.56)	3.7 (3.3–4.0)	0.00 (0.00–1.19)	2.92 (1.34–5.47)
Prepared for action—self-evaluative		0.79	2.94 (0.60)	3.0 (2.5–3.25)	0.00 (0.00–1.19)	0.32 (0.01–1.80)
Prepared for action—behavioural		0.52	3.86 (0.78)	4.0 (3.3–4.3)	1.30 (0.35–3.29)	9.74 (6.67–13.61)
Uncertain maintenance	66	0.82	3.27 (0.68)	3.2 (2.8–3.6)	0.00 (0.00–5.44)	0.00 (0.00–5.44)
Proactive maintenance		0.77	4.18 (0.65)	4.3 (4.0–4.5)	1.52 (0.04–8.16)	12.12 (5.38–22.49)
For hospital-recruited only						
Pre-contemplation	73	0.75	1.72 (0.88)	1.3 (1.0–2.3)	41.10 (29.71–53.23)	0.00 (0.00–4.93)
Contemplation		0.45	3.62 (0.55)	3.7 (3.3–4.0)	0.00 (0.00–4.93)	5.48 (1.51–13.44)
Prepared for action—self-evaluative		0.71	2.64 (0.59)	2.5 (2.3–3.0)	1.37 (0.03–7.40)	0.00 (0.00–4.93)
Prepared for action—behavioural		0.58	3.57 (0.84)	3.7 (3.0–4.3)	0.00 (0.00–4.93)	6.85 (2.26–15.26)
Uncertain maintenance	24	0.71	3.35 (0.74)	3.7 (2.9–3.8)	4.17 (0.11–21.12)	0.00 (0.00–14.25)
Proactive maintenance		0.89	4.01 (0.81)	4.3 (3.6–4.4)	4.17 (0.11–21.12)	12.50 (2.66–32.36)
Norwegian scoring model						
RTW inability	381	0.67	2.11 (0.69)	2.0 (1.6–2.6)	4.20 (2.42–6.73)	0.00 (0.00–0.96)
RTW uncertatinty		0.66	3.52 (0.61)	3.7 (3.0–4.0)	0.26 (0.01–1.45)	2.62 (1.27–4.77)
Uncertain maintenance	90	0.71	3.81 (0.70)	3.9 (3.4–4.2)	1.11 (0.03–6.04)	2.22 (0.27–7.80)
Proactive maintenance		0.49	3.41 (0.70)	3.6 (3.0–4.0)	0.00 (0.00–4.02)	2.22 (0.27–7.80)

Subscales scored from 1–5

High scores indicate high level of agreement

SD standard deviation, *iqr* interquartile range, *CI* confidence interval

Table 4 Model fit evaluated by CFI, TLI and RMSEA, SRMR

	χ^2	p	CFI	TLI	RMSEA	SRMR
Model fit for four stages for non-working (CA), n = 373	397.1	< 0.001	**0.934**	**0.913**	0.124	0.100
Model fit for two stages for non-working (NO), n = 377	222.8	< 0.001	**0.916**	0.877	0.169	0.124

Bold indicates indicative of acceptable model fit

CFI Comparative Fit Index, *TLI* Tucker-Lewis Index, *RMSEA* root mean square error of approximation, *SRMR* standardised root mean square residual

In the original RRTW-CA instrument, the internal consistency reliability of the stages was appraised as good; however, with only two stages (*prepared for action—self-evaluative* α = 0.75 *and uncertain maintenance* α = 0.82) [17] having Cronbach's alpha ranging between 0.70 and 0.95, the RRTW-DK instrument performed better; four out of six stages were within this recommended range.

Table 5 The correlation (Spearman's rho) between the stages of RRTW-DK and 19-item return-to-work self-efficacy (RTWSE-19)

RRTW-DK	n	rho	p value*
Total sample			
Pre-contemplation	355	−0.39	< 0.0001
Contemplation		0.18	< 0.001
Prepared for action self-evaluative		0.50	< 0.0001
Prepared for action behavioural		0.34	< 0.0001
Uncertain maintenance	81	−0.50	< 0.0001
Proactive maintenance		0.30	0.01
Job agency-recruited			
Pre-contemplation	291	−0.38	< 0.0001
Contemplation		0.12	0.04
Prepared for action self-evaluative		0.53	< 0.0001
Prepared for action behavioural		0.32	< 0.0001
Uncertain maintenance	60	−0.49	0.0001
Proactive maintenance		0.33	0.009

*Test for rho equals zero

The Norwegian scoring model did not show the same results as the original Canadian scoring model concerning stage differentiation and number of stages [8, 22]. In the present study, it could not be confirmed that the Norwegian scoring model had better internal consistency reliability than the Canadian scoring model as only one stage had a satisfactory alpha compared with four stages in the Canadian scoring model when applied on Danish data. After close inspection of the translation of the items, we found that the translated wording of A11 was the opposite of what was intended. Consequently, we reversed the scoring of the Danish A11 item so that a low/high score reflected a low/high degree of *contemplation* in both instruments. However, even after that, the alpha remained very low ($\alpha = 0.44$). Very poor correlations point to different strategies in how you begin to think about work resumption; either you handle things yourself (item A9) or you want to get help from others (item A11 and A12). This was confirmed in the Norwegian scoring model, where item A9 was left out [22].

For the stage *pre-contemplation*, 51% scored the lowest possible. However, all of those would have been allocated to higher stages if the stage allocation approach had been used. Neither Braathen [22] nor Franche [17] and Park [23] reported possible floor or ceiling effects in their evaluations of psychometric properties.

Structural Validity

The present study could not confirm the original Canadian scoring model; nor could Braathen [22] and Park [23]; both studies went along and performed exploratory factor analyses. We anticipated a better match between the Norwegian scoring model and the Danish data than between the Canadian scoring model and the Danish data. However, the RRTW-DK instrument reached acceptable fit estimates with the Canadian scoring model, whereas this was not the case with the Norwegian scoring model. We did not try to fit the RRTW-DK instrument to the scoring model found by Park et al. [23]. In spite of comparable outpatient rehabilitation settings, Braathen and Park obtained different scoring models; Park et al. found that disability duration, occupation and age were significantly associated with the RRTW stages, and the samples differed in particular with regards to disability duration [22, 23]. In the present study, the population's disability duration was unknown, and occupation was difficult to compare across the three studies. Age seemed to be comparable across the three studies.

The working sample was too small for CFA, but since the structural validity could not be confirmed in the non-working sample, we did not expect good fit indices on data from the working sample.

Construct Validity

In the original RRTW-CA instrument, construct validity was supported by correlations in the hypothesized direction with depressive symptoms, fear-avoidance beliefs, pain and general health [17]. Moreover, the concept of behaviour change has been thoroughly researched and received strong empirical support across 48 different health-related behaviours [35]. However, we find that the construct validity is questionable as two independent studies [22, 23] and the present study were unable to confirm the factor structure of the original RRTW-CA instrument as well as the direction of item scoring. Furthermore, correlations between RRTW-DK stages and the RTWSE-19 [24] did not consistently reach the threshold indicative of construct validity [33].

Methodological Strengths and Limitations

A substantial dropout rate between T0 and T1 limits generalization as responders were less likely than non-responders to suffer from chronic disease.

The minimum sample size of 50 individuals in validation studies recommended by Terwee et al. [20] was met; the statistical power was therefore sufficient to allow inferences from the results concerning test–retest reliability and SEM for the non-working sample, internal consistency reliability and the construct validity. In the confirmatory factor analyses, inferences could be made only based on the non-working sample.

Implications

It was not possible to confirm the original Canadian scoring model with the present data from the RRTW-DK; therefore, the RRTW-DK instrument in its current form cannot be recommended used.

Rather than keep performing explorative factor analyses in future cross-cultural validation studies, we recommend exploration of the construct in a homogenous population with regard to cause of sickness, disability duration and age as this may add knowledge to the RRTW construct.

Acknowledgements We want to thank Jens Laurids Jensen, Elin Sonne and Helle Holm Marcussen, Jobkompagniet, Silkeborg Municipality for their interest and involvement in this study. Last but not least we are grateful to all participants who kindly used their time answering questionnaires.

Funding We kindly thank The Danish Working Environment Research Fund, which funded this project (20-2013-09).

Compliance with Ethical Standards

Conflict of interest Author Christina Malmose Stapelfeldt, Anne-Mette Hedeager Momsen, Thomas Lund, Therese Koops Grønborg, Sheilah Hogg-Johnson, Janne Skakon, Chris Jensen and Merete Labriola declare that they have no conflict of interest.

Ethical Approval All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Approval for the use of questionnaire data was obtained from the Central Region Denmark (Danish Data Protection Agency j. no. 1-16-02-404-14). According to Danish law, approval from the Danish National Committee on Biomedical Research Ethics (http://www.cvk.sum.dk) was not relevant as approval is required only for projects using biological material or involving biomedical treatment. Participation in the study was voluntary, and answers were processed anonymously.

Appendix

The Readiness for Return to Work Instrument

For all items: Strongly disagree (1), Disagree (2), Neither disagree nor agree (3), Agree (4), Strongly agree (5).

For the items A7, A11 and B8, the scale is reversed.

Part (A) Not working sample

1. "You don't think you will ever be able to go back to work" *

2. "As far as you're concerned, there is no point in thinking about returning to work" * §

3. "You are actively doing things now to get back to work" *

4. "Physically, you are starting to feel ready to go back to work" *

5. "You have been increasing your activities at home in order to build up your strength to go back to work" *

6. "You are getting help from others to return to work"

7. "You are not ready to go back to work" *

8. "You have found strategies to make your work manageable so you can return to work" *

9. "You have been wondering if there is something you could do to return to work"

10. "You have a date for your first day back at work" *

11. "You wish you had more ideas about how to get back to work" * §

12. "You'd like to have some advice about how to go back to work" *

13. "As far as you are concerned, you don't need to go back to work ever" *

* Items with difficulties in the translation process

§ Items adjusted after pre-test in the adapted version of the questionnaire (RRTW-DK)

Stages in Part A Pre-contemplation (PC) A1, A2, A13

Contemplation (C) A9, A11, A12

Prepared for action self-evaluative (PA-S) A4, A7, A8, A10

Prepared for action behavioural (PA-B) A3, A5, A6

Part (B) Working Sample

1. "You are doing everything you can to stay at work"

2. "You have learned different ways to cope with your pain so that you can stay at work" *

3. "You are taking steps to prevent having to go off job again due to your injury" *

4. "You have found strategies to make your work manageable so you can stay at work"

5. "You are back at work but are not sure you can keep up the effort" *

6. "You worry about having to stop working again due to your injury"

7. "You still find yourself struggling to stay at work due to the effects of your injury"

8. "You are back at work and it is going well" *

9. "You feel you may need help in order to stay at work"

* Items with difficulties in the translation process

§ Items adjusted after pre-test in the adapted version of the questionnaire (RRTW-DK)

Stages in Part B Uncertain maintenance, items: B5, B6, B7, B8, B9

Proactive maintenance, items: B1, B2, B3, B4

References

1. Black DC, Frost D. Health at work—an independent review of sickness absence. 1st ed. London: Open Government Licence; 2011.
2. Henderson M, Glozier N, Holland Elliott K. Long term sickness absence. BMJ 2005;330(7495):802–803.
3. Palmer KT, Harris EC, Linaker C, Barker M, Lawrence W, Cooper C, et al. Effectiveness of community- and workplace-based interventions to manage musculoskeletal-related sickness absence and job loss: a systematic review. Rheumatology 2012;51(2):230–242.
4. Nieuwenhuijsen K, Faber B, Verbeek JH, Neumeyer-Gromen A, Hees HL, Verhoeven AC, et al. Interventions to improve return to work in depressed people. Cochrane Database Syst Rev. 2014;12:CD006237.
5. Nieuwenhuijsen K, Franche RL, van Dijk FJ. Work functioning measurement: tools for occupational mental health research. J Occup Environ Med. 2010;52(8):778–790.
6. de Boer AG, Taskila TK, Tamminga SJ, Feuerstein M, Frings-Dresen MH, Verbeek JH. Interventions to enhance return-to-work for cancer patients. Cochrane Database Syst Rev. 2015;9:CD007569.
7. Duijts S, Dalton SO, Lundh MH, Horsboel TA, Johansen C. Cancer survivors and return to work: current knowledge and future research. Psychooncology 2017;26(5):715–717.
8. Franche RL, Krause N. Readiness for return to work following injury or illness: conceptualizing the interpersonal impact of health care, workplace, and insurance factors. J Occup Rehabil. 2002;12(4):233–256.
9. Krause N, Frank JW, Dasinger LK, Sullivan TJ, Sinclair SJ. Determinants of duration of disability and return-to-work after work-related injury and illness: challenges for future research. Am J Ind Med. 2001;40(4):464–484.
10. Frank JW, Brooker AS, DeMaio SE, Kerr MS, Maetze lA, Shannon HS, et al. Disability resulting from occupational low back pain. part II: What do we know about secondary prevention? A review of the scientific evidence on prevention after disability begins. Spine (Phila Pa 1976). 1996;21(24):2918–2929.
11. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol. 1983;51(3):390–395.
12. Prochaska JO, DiClemente CC, Velicer WF, Ginpil S, Norcross JC. Predicting change in smoking status for self-changers. Addict Behav. 1985;10(4):395–406.
13. Prochaska JO, Velicer WF, DiClemente CC, Fava J. Measuring processes of change: applications to the cessation of smoking. J Consult Clin Psychol. 1988;56(4):520–528.
14. Prochaska JO, DiClemente CC, Velicer WF, Rossi JS. Criticisms and concerns of the transtheoretical model in light of recent research. Br J Addict. 1992;87(6):825–828 (**discussion 833–835**).
15. Prochaska JO, Diclemente CC, Norcross JC. In search of how people change. applications to addictive behaviors. Am Psychol. 1992;47(9):1102–1114.
16. Prochaska JO, Velicer WF, Rossi JS, Goldstein MG, Marcus BH, Rakowski W, et al. Stages of change and decisional balance for 12 problem behaviors. Health Psychol. 1994;13(1):39–46.
17. Franche RL, Corbiere M, Lee H, Breslin FC, Hepburn CG. The readiness for return-to-work (RRTW) scale: development and validation of a self-report staging scale in lost-time claimants with musculoskeletal disorders. J Occup Rehabil. 2007;17(3):450–472.
18. Bandura A. Self-efficacy: the exercise of control. New York: W.H. Freemann and Company; 1997.
19. Krause N, Ragland DR. Occupational disability due to low back pain: a new interdisciplinary classification based on a phase model of disability. Spine (Phila Pa 1976) 1994;19(9):1011–1020.
20. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42.
21. Amick BC III, Lerner D, Rogers WH, Rooney T, Katz JN. A review of health-related work outcome measures and their uses, and recommended measures. Spine (Phila Pa 1976) 2000;25(24):3152–3160.
22. Braathen TN, Brage S, Tellnes G, Eftedal M. Psychometric properties of the readiness for return to work scale in inpatient occupational rehabilitation in Norway. J Occup Rehabil. 2013;23(3):371–380.
23. Park J, Roberts MR, Esmail S, Rayani F, Norris CM, Gross DP. Validation of the readiness for return-to-work scale in outpatient occupational rehabilitation in Canada. J Occup Rehabil. 2018;28(2):332–345.
24. Hedeager Momsen AM, Rosbjerg R, Stapelfeldt CM, Lund T, Jensen C, Johansen T, et al. Cross-cultural adaptation and validation of the danish version of the 19-item return-to-work self-efficacy (RTWSE-19) questionnaire. Scand J Work Environ Health. 2016;42(4):338–345.
25. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). 2000;25(24):3186–3191.
26. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international delphi study. Qual Life Res. 2010;19(4):539–549.
27. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010;10(1):1–8.
28. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737–745.
29. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995;4(4):293–307.
30. Bland JM, Altman DG. Measurement error and correlation coefficients. BMJ 1996;313(7048):41–42.
31. Nunnally J. Psychometric theory. 2th ed. New York: McGraw-Hill; 1978. p. 245.
32. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Model. 1999;6(1):1–55.
33. Hemphill JF. Interpreting the magnitudes of correlation coefficients. Am Psychol. 2003;58(1):78–79.
34. Beta-version of the readiness for return to work [parathed.pdf] [homepage on the Internet]. Aarhus, Denmark: DEFACTUM. 2017. http://bit.ly/2B01SRC. Accessed 15 April 2017
35. Hall KL, Rossi JS. Meta-analytic examination of the strong and weak principles across 48 health behaviors. Prev Med. 2008;46(3):266–274.