Original Empirical Research

Psychometric properties of two implementation measures: Normalization MeAsure Development questionnaire (NoMAD) and organizational readiness for implementing change (ORIC)

P. Batterham[1], Caroline Allenhof[2] (ID), Arlinda Cerga Pashoja[3,4], A. Etzelmueller[5,6], N. Fanaj[7], T. Finch[8], J. Freund[9,10], D. Hanssen[11], K. Mathiasen[12,13], J. Piera-Jiménez[14,15,16], G. Qirjako[17,18], T. Rapley[19], Y. Sacco[20], L. Samalin[21], J. Schuurmans[22], Claire van Genugten[23,24] and C. Vis[24,25,26] (ID)

Abstract

Background

Effective interventions need to be implemented successfully to achieve impact. Two theory-based measures exist for measuring the effectiveness of implementation strategies and monitor implementation progress. The

[1]Centre for Mental Health Research, Australian National University, Canberra, Australia
[2]German Foundation for Research and Education on Depression, Leipzig, Germany
[3]London School of Hygiene & Tropical Medicine, London, UK
[4]St. Marys University Twickenham, UK
[5]HelloBetter, GET.ON Institut für Online Gesundheitstrainings GmbH, Hamburg, Germany
[6]Department Health and Sport Sciences, Technical University of Munich, School of Medicine and Health, Professorship Psychology & Digital Mental Health Care, München, Germany
[7]Alma Mater Europaea Campus College Rezonanca, Pristina, Kosovo
[8]Department of Nursing, Midwifery & Health, Northumbria University, Newcastle upon Tyne, UK
[9]Department of Clinical Psychology and Psychotherapy, Institute of Psychology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[10]Professorship Psychology and Digital Mental Health Care, Department of Sport and Health Sciences, Technical University of Munich, Munich, Germany
[11]Interdisciplinary Centre Psychopathology and Emotion Regulation, University Medical Centre Groningen, Groningen, Netherlands
[12]Department of Clinical Research, University of Southern Denmark, Odense, Denmark
[13]Centre for Digital Psychiatry, Lillebaelt Hospital – University Hospital of Southern Denmark, Vejle, Denmark
[14]Government of Catalonia Catalan Health Service, Barcelona, Spain
[15]Digitalization for the Sustainability of the Healthcare System (DS3), Institut d'Investigacions Biomèdiques de Bellvitge (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain
[16]Faculty of Informatics, Telecommunications and Multimedia, Universitat Oberta de Catalunya, Barcelona, Spain

[17]Department of Public Health, University of Medicine of Tirana, Tirane, Albania
[18]Community Centre for Health and Wellbeing, Tirane, Albania
[19]Department of Social Work, Education and Community Wellbeing, Northumbria University, Newcastle upon Tyne, UK
[20]Fondazione Don Carlo Gnocchi, Presidio Ausiliatrice S. Maria ai Colli, Torino, Italy
[21]Department of Psychiatry, CHU Clermont-Ferrand, University of Clermont Auvergne, CNRS, Clermont Auvergne INP, Institut Pascal (UMR 6602), Clermont-Ferrand, France
[22]Taytelbaum Psychologen, Amsterdam
[23]Clinical, Neuro-, and Developmental Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[24]Mental Health, Amsterdam Public Health Research Institute, Amsterdam, Netherlands
[25]Public and Occupational Health, Amsterdam University Medical Center, Amsterdam, Netherlands
[26]Forhelse Research Centre for Digital Mental Health Services Division of Psychiatry, Haukeland University Hospital, Bergen, Norway

Corresponding author:
C. Vis, Public and Occupational Health, Amsterdam University Medical Centres, Van der Boechorststraat 7, Amsterdam, Noord-Holland 1083BA, The Netherlands.
Email: p.d.c.vis@amsterdammumc.nl

Normalization MeAsure Development questionnaire (NoMAD) explores the four core concepts (Coherence, Cognitive Participation, Collective Action, Reflexive Monitoring) of the Normalization Process Theory. The Organizational Readiness for Implementing Change (ORIC) is based on the theory of Organizational Readiness for Change, measuring organization members' psychological and behavioral preparedness for implementing a change. We examined the measurement properties of the NoMAD and ORIC in a multi-national implementation effectiveness study.

Method

Twelve mental health organizations in nine countries implemented Internet-based cognitive behavioral therapy (iCBT) for common mental disorders. Staff involved in iCBT service delivery ($n = 318$) participated in the study. Both measures were translated into eight languages using a standardized forward–backward translation procedure. Correlations between measures and subscales were estimated to examine convergent validity. The theoretical factor structures of the scales were tested using confirmatory factor analysis (CFA). Test–retest reliability was based on the correlation between scores at two time points 3 months apart. Internal consistency was assessed using Cronbach's alpha. Floor and ceiling effects were quantified using the proportion of zero and maximum scores.

Results

NoMAD and ORIC measure related but distinct latent constructs. The CFA showed that the use of a total score for each measure is appropriate. The theoretical subscales of the NoMAD had adequate internal consistency. The total scale had high internal consistency. The total ORIC scale and subscales demonstrated high internal consistency. Test–retest reliability was suboptimal for both measures and floor and ceiling effects were absent.

Conclusions

This study confirmed the psychometric properties of the NoMAD and ORIC in multi-national mental health care settings. While measuring on different but related aspects of implementation processes, the NoMAD and ORIC prove to be valid and reliable across different language settings.

Plain Language Summary: Why was the study done? Effective interventions need to be implemented successfully to achieve impact. Reliable measurement instruments are needed to determine if an implementation was successful or not. Two theory-based instruments exist for measuring the effectiveness of implementation strategies and monitor progress. The NoMAD measures aspects of normalization related to sense-making, willingness to implement, the work people do, and reflection. The Organizational Readiness for Implementing Change (ORIC) measures organization members' preparedness for implementing a change. **What did we do?** This study examined whether the NoMAD and ORIC measure what they are supposed to measure. We translated the instruments from English to eight languages (Albanian, Danish, Dutch, French, German, Italian, and Spanish/Catalan) We applied various statistical methods to confirm the measurement properties, including correlations of scales, factor structures, test–retest reliability, consistency and floor and ceiling effects. 318 mental health professionals from nine countries participated in the study. **What did we find?** For both instruments, total scores can be used as well as the subscale scores. Internal consistency for ORIC was high and for NoMAD adequate. Test–retest reliability was demonstrated, and floor and ceiling effects were rare. **What does this mean?** NoMAD and ORIC are reliable instruments for measuring implementation processes and outcomes across mental health care settings in different countries and languages. They measure related but different aspects of implementation processes and outcomes. The measures are brief, and theory supported. However, more work is to be done on interpreting scores in relation to implementation success and regarding changes over time.

Background

To achieve evidence-based practice, interventions that are proven to be effective need to be implemented successfully. Numerous studies have emerged investigating and reporting on factors that hinder or can facilitate implementation processes (Krause et al., 2014; Vis et al., 2018). A number of taxonomies and frameworks such as the Consolidated Framework for Implementation Research (CFIR; Damschroder et al., 2022) and Implementation Mapping (IM; Fernandez et al., 2019) have been developed to aid understanding of a particular health practice and guide the development of strategies that can improve the implementation of new interventions in that practice (Grimshaw et al., 2004; Nilsen, 2015; Proctor et al., 2013).

Valid and reliable measures are required to determine the effectiveness of implementation strategies and to monitor progress in implementing interventions into practice. As with any measure, implementation measures should be conceptually sound, valid, and reliable. However, as implementation science is an applied science, measurement instruments should also be relevant and usable in practice, that is, pragmatic measures balancing external validity with internal validity (Glasgow et al., 2006). Pragmatic measures emphasize the context in which they are (intended) to be used and enable people to address unique qualities of the intervention that is implemented to ensure the measurement contributes to improving practice in a meaningful way (Glasgow, 2013).

In recent years, many different measurement instruments have been developed and used in process evaluations of clinical interventions and implementation studies. Increasingly, choices about outcomes are structured and selected using frameworks such as proposed by RE-AIM (Glasgow et al., 1999) or Proctor et al. (2011), providing suggestions for relevant areas of measurement such as the acceptability and reach of the intervention, as well as perceived appropriateness and levels of adoption. These measurement domains subsequently need to be operationalized into items that are relevant for a particular implementation project.

Until recently, many measurement instruments that are developed and reported on are pragmatic, intervention, study and setting specific, and lack a theoretical basis (Martinez et al., 2014). We know that theory and theory-based measurement instruments are essential in formulating and testing causal relations between intervention (components) and observation, can aid in interpreting findings, and ultimately, provide an empirical basis for recommending or rejecting certain implementation strategies because of their level of effectiveness. We also know that it is essential in establishing the validity and reliably of measurement instruments through psychometric assessment to ensure what is measured is an actual representation of real practice.

Only recently, work in this area has advanced for aspects measuring implementation readiness including measures of leadership, available resources, and knowledge and information (Weiner et al., 2017, 2020a, 2020b). Generally, validated theory-based instruments to objectively and reliably monitor implementation processes and assess outcomes are lacking (Lewis et al., 2015a, 2015b; Mettert et al., 2020; Proctor et al., 2011; Weiner et al., 2020a, 2020b).

Implementation theories such as Normalization Process Theory (NPT; May et al., 2009) and Organizational Readiness for Change (ORC; Weiner, 2009) can provide heuristic tools to aid the development of practical and sensible measurement instruments. Both theories have good face validity (May et al., 2018; Weiner et al., 2020a, 2020b) and are perceived as useful in explaining and hypothesizing implementation processes and outcomes. NPT explains the work people do when integrating and embedding new ways of working in current practices.

ORC, on the other hand, places the concept of readiness as central. This is determined by peoples' change valence and change efficacy through which a change is implemented. Both theories draw on several sociological and organizational theories including motivation theory (Fishbein & Ajzen, 1977; Meyer & Herscovitch, 2001; Vroom, 1964) and social cognitive theory (Gist & Mitchell, 1992), and are initially based on qualitative studies of implementation processes in health care.

For both theories, pragmatic measurement instruments were developed. The Normalization MeAsure Development project (NoMAD; Finch et al., 2013) is a brief self-report questionnaire that focuses on the four core concepts of the NPT. The Organizational Readiness for Implementing Change (ORIC; Shea et al., 2014) measures organization members' psychological and behavioral preparedness for implementing organizational change. Although basic psychometric properties have been tested for these instruments, others remain unknown. Notably, psychometric evidence of their construct validity, test–retest stability, and floor and ceiling effects can help refine the understanding of the qualities of the NoMAD and ORIC as measures for implementation readiness and success (Terwee et al., 2007; Vis et al., 2019). Furthermore, the conceptual interpretation of the scores requires further explanation to enable targeted interventions based on process and outcome assessments and comparisons and increase practical utility of these measures. Besides expansion on these psychometric properties, the validity of the measures across (national) health systems and settings and languages has not been established. Therefore, this study aimed to test examine the psychometric properties of the NoMAD and ORIC in multiple mental health settings in different countries using different languages.

Method

Following quality criteria for measurement instruments (Terwee et al., 2007), a psychometric study was conducted to examine the internal consistency, construct validity, convergent validity, and reproducibility (test–retest reliability; floor/ceiling effects).

The current psychometric study concerned a post-hoc secondary analysis of data collected in a multi-national, multi-site implementation study (Buhrmann et al., 2020; Vis et al., 2023) testing the effectiveness of tailored implementation of evidence-based digital mental health interventions for mental health disorders (depression, anxiety, somatic symptoms disorder). Tailoring involves systematically identifying factors hindering and facilitating implementation, designing implementation interventions appropriate to those determinants, and applying those interventions (Baker et al., 2015; McHugh et al., 2022; Wensing, 2017). A self-guided online toolkit (www. ItFits-toolkit.com) was developed that provides implementation teams with a stepped and structured working process to identify objectives and barriers, select one or more

implementation strategies and to develop, apply, and evaluate an actionable work plan. The toolkit was tested for effectiveness in a stepped-wedge cluster randomized controlled trial in 12 mental health care organizations in nine countries (Buhrmann et al., 2020). The ItFits-toolkit had a small significant positive effect on normalization levels in service delivery staff compared to usual implementation activities (Vis et al., 2023).

NoMAD

NoMAD was developed to measure constructs of NPT. NPT starts from the proposition that normalizing (i.e., embedding and integrating) innovation in routine practice is a result of the things people individually and collectively do. According to NPT, normalization processes are driven by four generative mechanisms: (a) coherence of the innovation with the (goals of) daily routine, (b) cognitive participation as a process of enrollment and engagement of individual participants and groups, (c) collective action by individuals and groups to apply the innovation in daily routines, and (d) reflexive monitoring through which participants in the implementation process evaluate and appraise the use of the new intervention in practice.

NoMAD was designed to aid identifying normalization specific barriers, to monitor progress of implementation projects in practice, and to compare normalization outcomes between implementation projects and over time. The NoMAD questionnaire in this study consisted of 23 items. The first three items concerned general normalization ratings about the current use and likelihood of using the intervention in the future. These items were scored on a 1–10 Visual Analog Scale (VAS). Following the original validation study, these items were added as control questions to assess its convergent validity. The remaining 20 items were allocated to four subscales to measure the four NPT constructs: Coherence (CO) with four items; Cognitive Participation (CP) with four items; Collective Action (CA) with seven items; and Reflexive Monitoring (RM) with five items.

NoMAD underwent a thorough development process (Rapley et al., 2018) and its conceptual and internal validity have been studied and confirmed in various settings (Elf et al., 2018; Vis et al., 2019; Finch et al., 2018; Jiang et al., 2022; Loch et al., 2020; Freund et al., 2023). Confirmatory factor analysis (CFA) showed the model to approximate the data acceptably. Construct validity of the four theoretical constructs of NPT was supported, and internal consistency for the global sum score was high and for the four subscales acceptable. The overall normalization scale was found to be a reliable measure for the concepts of interest.

ORIC

ORIC was designed to measure the two constructs of Change Commitment (CC) and Change Efficacy (CE) as defined by the conceptual framework by Weiner (Weiner, 2009) for Organizational Readiness for Change. CC refers to the shared resolve of organizational members to implement an intervention. CE is determined by change valence which refers to the (shared) value organizational members' attribute to the new intervention and the required change attached to it. Change efficacy refers to organizational members' shared belief in their collective capabilities to implement the intervention. Change efficacy is determined by members' knowledge of what is required of them, the availability of required resources, and other situational factors such as time and support from the management. In essence, change efficacy concerns an informational element of implementing a change.

ORIC was developed using four laboratory and empirical studies (Shea et al., 2014). The final measure consists of 12 items: five items addressing change commitment and seven items for change efficacy. The conceptual and internal validity have been studied and the questionnaire showed good model fit, high item loadings, and inter-item consistency for the two subscales (Shea et al., 2014).

The primary target group for both measurement instruments is staff members involved in implementing change. Both ORIC and NoMAD are rated on a 5-point Likert scale ($1 = I$ *disagree* to $5 = I$ *agree*). To prevent ambiguity and reduce missingness, a forced-choice approach was applied for both questionnaires. Respondents were required to rate all statements. Sub- and total scale scores were calculated by taking the mean of answered items of a scale. For both measures, a higher score was interpreted as better normalization (NoMAD) and readiness for implementing change (ORIC).

Demographics and Translations

Besides the NoMAD and ORIC items, basic demographic information including gender, age, years of relevant working experience, professional job category, and relevant care sector were collected. The English NoMAD, ORIC, and demographic items are included in Appendix A. A forward–backward translation procedure was used to translate the original English NoMAD and ORIC questionnaires into six languages (Albanian, Danish, French, German, Italian, and Catalan). One translation (Dutch) was available from earlier work using the same translation procedure. Back translations were done by a third translator and were confirmed by the principal investigators of the English NoMAD and ORIC for their equivalence of semantic meaning of the corresponding individual items.

Settings

This psychometric study was part of the ImpleMentAll project in which twelve implementation sites (i.e., mental health service delivery organizations) in nine countries (Albania, Australia, Denmark, France, Germany, Italy, Kosovo, the Netherlands, and Spain), implemented an Internet-based cognitive behavioral therapy (iCBT) services for prevention and treatment of

common mental disorders. All services were based on proven therapeutic principles (i.e., psychoeducation, techniques invoking behavioral change, a cognitive restructuring component, and relapse prevention) and included tested Internet delivery formats. The specific operationalization of the iCBT service (e.g., form and guidance in delivery, treatment protocol, etc.) varied in response to the local requirements (Buhrmann et al., 2020). Services were implemented in a combination of public health settings ($n = 3$), primary care ($n = 5$) and specialized care settings ($n = 6$).

Sample and Recruitment

Staff involved in iCBT service delivery at the local implementation sites were recruited to participate in the study. Service delivery staff were eligible to participate when (a) they were involved in the delivery of the iCBT service, and (b) were in a process of adapting their current way of working in order to deliver their iCBT service to patients in routine care. Staff members could have different roles in the service delivery: as therapists, as referrers, as administrators or as ICT support staff. Recruitment of staff members started prior to the baseline measure. Participants were recruited by the implementation sites through information sessions, direct and general e-mailings, phone, and in-person depending on their involvement and local trial site characteristics (Buhrmann et al., 2020). Sample size calculations were based on a series of simulation studies using multi-level degree of normalization. With 15 staff participants per implementation site per measurement wave it was proven to provide >80% power to detect effectiveness (two-sided, $\alpha = .05$). Assuming a drop-out of 20%, the target sample size was set to $n = 19$ per implementation site resulting in 228 participants per data collection wave (Buhrmann et al., 2020). Staff respondents were replaced when not responding until the target sample size was reached. The most important reason for non-responding was staff turnover (other roles in organization, leaving organization).

Measurements and Procedures

The study lasted for 30 months and included ten data collection waves. Every 3 months the NoMAD and ORIC were issued online in local language to the included staff participants. For the purpose of this psychometric study, we used data from Wave 2 as the sample was larger than at Wave 1 when recruitment was still being conducted and occurred prior to the introduction of the ItFits toolkit. The present analyses included 318 participants who completed the Wave 2 NoMAD assessment, with slightly fewer ($n = 313$, 98%) completing the ORIC assessment. Data collected in Wave 3 was used for establishing the test–retest reliability. It must be noted that during Wave 3, the first two implementation sites were introduced to the ItFits toolkit following the stepped-wedge RCT design (Buhrmann et al., 2020). Although the measurement was opened prior to the introductory training of the implementation teams, it could be that some respondents in those sites filled out the NoMAD and ORIC after the introduction. This could not be more than a few days and it is highly unlikely that the introduction of the toolkit to the implementation teams had an impact on delivery staff in such short timeframe.

Local research teams managed the inclusion of participants and data collection procedures using a purpose-built central data collection system. Demographic information was collected at enrollment of the participant.

Analysis

Analyses were conducted using SPSS v29 (IBM Corp, Chicago, USA), and confirmatory factor analyses were conducted using Mplus v6.12 (Muthén & Muthén, Los Angeles, USA). Correlations between measures and subscales were estimated to examine convergent validity. Confirmatory factor analyses treated scale items as ordered categorical variables and used a robust weighted least squares estimator (Muthén, 1984). Because the scales are designed to be multidimensional, we tested multidimensional models (four-factor for the NoMAD, two-factor for the ORIC in correspondence with their respective theoretical basis), along with second-order models that assume a hierarchy with a single higher-order factor and specific lower-order factors (Mansolf & Reise, 2017). The fit of each scale to these structures was assessed on the basis of the root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker–Lewis index (TLI). Cutoffs for satisfactory and good fit were 0.90 and 0.95, respectively, for CFI and TLI, and 0.06 and 0.08, respectively, for RMSEA (Hu & Bentler, 1998). Where good fit was not demonstrated, we explored dropping items with high local dependence (based on modification indices) or by accounting for correlations between items with local dependence. Test–retest reliability was based on the correlation between scores at Wave 2 and Wave 3, conducted 3 months apart. The subsample who completed the NoMAD ($n = 280$, 88%) and ORIC ($n = 272$, 86%) was used in the estimation of test–retest reliability. We assumed good test–retest reliability when $r > .7$ (Terwee et al., 2007). Internal consistency was assessed using Cronbach's alpha where we assumed good consistency when Cronbach's alpha(s) are between .70 and .95 (Terwee et al., 2007). Floor and ceiling effects were quantified based on the proportion of the sample reporting a zero or maximum score respectively on each total scale. We assumed floor or ceiling effects were present if more than 15% of respondents achieved the lowest or highest possible score, respectively (Terwee et al., 2007). Only complete cases were used.

Ethics

This study was part of a multi-center implementation effectiveness trial. Non-medical personal data was collected and processed from human healthy volunteers. Participants were required to provide informed consent (in local language) prior to data collection explaining the purpose of the study,

nature, use, and management of their data. Local research teams translated the study protocol and applied for approval from competent local medical ethical review committees. All committees approved the study. A portfolio of the ethical documentation was issued to the European Commission (Deliverable 9.1, Horizon 2020 Grant Agreement no. 733025). The study protocol was published (Buhrmann et al., 2020) and registered with ClinicalTrials.gov (NCT03652883).

Results

Demographics and Scores

Among the 318 participants included in the analyses, 211 (66%) were women and 107 (34%) men. Participant ages ranged from 18 to 75 years, with a mean of 43.1 ($SD = 10.6$) and median of 43 years. Most participants had worked for less than 10 years in their organization, with 87 (27%) reporting 0–2 years, 115 (36%) 3–10 years, 55 (17%) 11–15 years, and 61 (19%) more than 10 years. Roles in commissioning the iCBT services were varied: a majority of participants were in a therapist role (e.g., psychologist, psychiatrist, mental health nurse: 171, 54%), 119 (37%) were a referrer (e.g., GP, pharmacist, community worker, case manager), 25 (8%) were administrators (e.g., clerical worker, secretarial) and three (1%) were in ICT support (e.g., security officer, maintenance officer, helpdesk staff). Most participants ($n = 240$, 76%) had no previous involvement or experience in delivering internet-based CBT programs.

Mean total scores for the NoMAD were 3.68 ($SD = 0.66$), reflecting an average response between "neither agree/disagree" and "agree" on the 5-point scale. Mean total scores for the ORIC were 3.54 ($SD = 0.82$), similarly reflecting an average response between "neither agree/disagree" and "somewhat agree" on the 5-point scale. There were no significant associations between demographic characteristics or experience/role with either scale, with two exceptions. NoMAD scores were significantly lower for referrers than

those in other roles ($F = 3.29$, $df = 3$, 314, $p = .021$). ORIC scores were significantly higher for participants with no previous experience in delivering iCBT ($F = 3.83$, $df = 1$, 311, $p = .016$) with higher scores indicating higher expression of the NoMAD/ORIC subscales.

Correlations between NoMAD and ORIC total scores and subscales are provided in Table 1. All correlations were moderate to strong, ranging from .51 to .98. Total scores on the two scales were moderately correlated ($r = .678$). Subscales within each scale were more highly correlated with other subscales from the same measure than with subscales from the other measure.

Factor Structure

The factor structures of the scales were tested in CFA models. The fit indices for the models are provided in Table 2. The multifactor models had inadequate fit for both measures, with the second-order models performing better. This indicates that the scales can be represented by both a higher-order factor (total score) and by subscales representing two (ORIC) and four (NoMAD) specific factors.

CFI and TLI indicated adequate good fit for the NoMAD second-order model. RMSEA was close to the .08 cutoff, with a confidence interval overlapping .08. This finding suggests the second-order model was adequate. Nevertheless, we explored whether excluding the item with most local dependence (highest modification indices) with other items would improve the fit. This item was item CA.2: "[The intervention] disrupts working relationship"; dropping the item led to only marginally improvement in fit.

CFI and TLI indicated good fit for the ORIC second-order model; however, RMSEA was below an adequate level. Consequently, we accounted for high local dependence between items CE.7 ("… confident that they can keep the momentum going in implementing the change") and CE.8 ("… confident that they can handle the challenges that might arise in implementing the change"). This led to better fit, with the RMSEA confidence interval overlapping .08. When removing item CE.7, fit was further improved.

Table 1

Correlations Between NoMAD and ORIC Total Scores and Subscales.

	NoMAD	CO	CP	CA	RM	ORIC	CC	CE
NoMAD total	1.000	0.846	0.849	0.898	0.897	0.678	0.663	0.654
Coherence (CO)	0.846	1.000	0.686	0.640	0.712	0.536	0.527	0.514
Cognitive participation (CP)	0.849	0.686	1.000	0.661	0.692	0.572	0.551	0.558
Collective action (CA)	0.898	0.640	0.661	1.000	0.732	0.615	0.607	0.589
Reflexive monitoring (RM)	0.897	0.712	0.692	0.732	1.000	0.630	0.611	0.610
ORIC global score	0.678	0.536	0.572	0.615	0.630	1.000	0.956	0.979
Change commitment (CC)	0.663	0.527	0.551	0.607	0.611	0.956	1.000	0.876
Change efficacy (CE)	0.654	0.514	0.558	0.589	0.610	0.979	0.876	1.000

Note. NoMAD = Normalization MeAsure Development Scale; ORIC = Organizational Readiness for Implementing Change Scale. All correlation coefficients have $p < .001$.

Table 2
Fit Indices for Confirmatory Factor Analysis Models.

	RMSEA	90% CI	CFI	TLI
NoMAD				
Four-factor model	0.102	0.095, 0.110	**0.920**	**0.908**
Second-order model	0.083	**0.075, 0.092**	**0.954**	**0.939**
Second-order model excluding item CA.2	0.081	**0.072, 0.090**	**0.960**	**0.947**
ORIC				
Two-factor model	0.125	0.112, 0.138	**0.983**	**0.979**
Second-order model	0.106	0.090, 0.121	**0.991**	**0.985**
Second-order model + correlated items	0.086	**0.070, 0.102**	**0.994**	**0.990**
Second-order model excluding item CE.7	**0.071**	**0.053, 0.090**	**0.996**	**0.994**

Note. RMSEA = root-mean-square error of approximation; CI = confidence interval; CFI = comparative fit index; TLI = Tucker–Lewis index; NoMAD = Normalization MeAsure Development Scale; ORIC = Organizational Readiness for Implementing Change Scale. Bold values indicate acceptable fit.

Reliability

The subscales of the NoMAD had adequate internal consistency: Cronbach's α = .761 for coherence, .808 for cognitive participation, .781 for collective action and .773 for reflexive monitoring. The total scale had high internal consistency, with α = .921. Test–retest reliability was suboptimal based on the correlations between scores at Wave 2 and Wave 3, with r = .674 for the total score.

The full ORIC scale and subscales demonstrated high internal consistency: Cronbach's α = .907 for change commitment, .940 for change efficacy and .959 for the total scale, noting that high alpha (>.95) may be indicative of redundancy in the items. Test–retest reliability was suboptimal for ORIC scores, with a correlation of r = .639 for total scale scores at Wave 2 versus Wave 3. Potential contamination by introduction of the toolkit in Wave 3 is negligible as the measurements were issued prior to the introduction of the toolkit to implementation teams in the first two sites and because of the large intervention-lag effect. Floor and ceiling effects were not present as there were no participants with a minimum possible score on the NoMAD, one participant (0.3%) with a maximum score on the NoMAD, two participants (0.6%) with a minimum score on the ORIC and nine participants (2.9%) with a maximum ORIC score.

Discussion

The implementation outcome measures NoMAD and ORIC were psychometrically tested across multiple languages and health care contexts. The latent conceptual structures were confirmed indicating acceptable construct validity. Both measures also converged to related but distinct (latent) factors. The adequate good fit of the second-order (hierarchical) models also indicates that both total scores and subscale scores are meaningful for both measures, representing a general factor with a series of specific factors representing the subscales. Reproducibility and internal consistency proved to be acceptable but can be improved. Some of the limitations of the psychometric properties may be related to the diversity of items, some of which capture current implementation efforts, while others address intentions for future implementation activity. Therefore, higher scores may reflect a perception of either greater implementation progress or of greater potential for positive implementation outcomes. Further exploration of how respondents interpret the items may provide further insights into the meaning of increases or decreases in scores on these measures, particularly in the context of implementation trials.

For the NoMAD, findings were similar to other studies in a variety of settings and languages: English nursing, sport and behavioral health care (Finch et al., 2018), Chinese nursing and injury health care (Jiang et al., 2022), Brazilian Portuguese HIV/AIDS clinical monitoring (Loch et al., 2020), Swedish post-discharge collaborative care planning (Elf et al., 2018), German prevention, primary care, and mental health care settings (Freund et al., 2023), and Dutch specialized and primary mental health care (Vis et al., 2019). Although all studies found acceptable internal consistency and factor structures, most suggested changing or omitting specific items. For example, the Swedish study (Elf et al., 2018) suggested deleting items related to the constructs Collective Action (CA.2, CA.3, CA.4) and Reflexive Monitoring (RM.4) to achieve acceptable model fit. The Dutch study (Vis et al., 2019) identified the same item (CA.2) scoring below cutoff and error terms to be added for CA.3 and CA.4 to improve model fit. In the German study, RM.1 was removed leading to a slightly better fit than the previous model. It must be noted that these changes are proposed based on empirical statistical associations. Construction of and changes to measurement instruments need to conceptually make sense and have solid evidence that the benefits of changing a validated measure outweigh the risks, such as problems with inconsistent scale versions. Given these findings, we suggest reconsidering item CA.2 "the intervention disrupts working relationships," which may involve including clarifications of the term "disruption" and "working relationships." Interpretation of this item may vary with the role and responsibilities respondents perceive to have in their organization in implementing the intervention. For example, a psychotherapist might more commonly operate alone working with one patient

in a session, whereas a surgeon may collaborate with a team of professionals during a medical procedure. Disruption and the impact of the intervention on working relations may manifest differently and therefore, can have a different meaning in relation to implementing a new practice. As such, contextualizing the questionnaire in the specific setting and context it will be used, is recommended to improve conceptual fit. It must be noted, however, that using slightly varied versions of the NoMAD may create limitations in generalizing conclusions about implementation success across varied settings.

For ORIC, similar observations can be made. In English-speaking lab test with students in health policy and a field test with staff of international non-governmental organizations involved in monitoring and evaluation of health programs (Shea et al., 2014), it was found that a model fit was improved by excluding two items for change efficacy (CE.1 and CE.4 related to the maintain momentum and engage people in the change, respectively). While most consecutive studies have used the adapted version with good psychometric results (Lindig et al., 2020; Ruest et al., 2019), only a few used the full measurement model allowing for replication of the original study. For example, a Danish study in obstetrics and gynecology in a university medical center found that a good fit was obtained when removing only item CE.1 (Storkholm et al., 2018). In Brazilian primary health care settings, also item CE.1 was removed after an exploratory factor analysis (Bomfim et al., 2020). In our study (also applying the original measurement model), we found that a better model fit was achieved when removing CE.7. One explanation could be that this item can be interpreted with notions of motivation that can be related to commitment rather than efficacy, leading to noise in the measurement model. Moreover, different items have been identified as reducing model fit across studies, suggesting that there may be redundancy or variable interpretation of concepts related to organizational readiness depending on the setting. As with the NoMAD, it may be appropriate to customize items to fit the implementation setting, noting that this customization may limit comparability across studies.

Strengths and Limitations

This study examined the psychometric properties of two key implementation measures, in a large sample of staff working in diverse health settings across multiple nations and languages. We used CFA to examine the fit of the scales to multidimensional and second-order models and examined important indicators of psychometric performance. However, there were some limitations. Our estimates of test–retest reliability are likely to be conservative given that the period between data collection waves was greater than typically used to assess this metric, which may explain why these estimates did not meet the threshold. Given the number of translations used, we did not have a sufficient sample in any single language to examine

linguistic invariance. The diversity of the sample was both a strength and weakness of the study—it is likely that the psychometric properties of the scales would be strengthened by examining response patterns in a more homogeneous sample. We were unable to test sensitivity to change in the context of the complex (stepped-wedge) trial design. The finding that participants tended to respond around the midpoint of the scales may suggest that some participants may have viewed certain items as less relevant in their implementation setting. Further tailoring of the scales to ensure broad applicability may lead to greater variability in responses. Finally, although stand-alone psychometric studies are unlikely to be warranted, further examination of scale properties within future implementation trials may provide additional insights into the validity, utility and notably interpretability of the measures. Such research might compare the scales to other indices of implementation success such as adoption.

Conclusions

The NoMAD and ORIC had acceptable measurement properties in the context of a multi-national implementation study. However, CFA indicated that some items of the scales may be redundant or require further clarification that may be tailored to the implementation setting. Despite these concerns, it remains imperative to identify the processes by which implementation succeeds within health settings, and both the NoMAD and ORIC are likely to provide important insights into the implementation of evidence-based interventions into practice. Future research may examine the meanings of high or low scores across different settings and identify whether targeted strategies to address different features of normalization or organizational readiness cause differential change across the subscales of the NoMAD and ORIC. Further exploration of the trade-off between maintaining standardization and adaptation of items for specific settings may also be warranted.

Authors' Contributions

CV and JP designed this study. Note this was part of a larger study. TF, TR, CS, JS, CRV, KM, JPJ and CV were involved in design of the measures, central study coordination. PB, CA, CPC, AE, NF, JF, DH, KM, JPJ, GQ, YS, LS, JS, CRG collected the data. PB and CV performed the analysis and wrote the manuscript. All coauthors reviewed, provided feedback, and approved publication.

Declaration of Conflicting Interests

Funding

ORCID iDs

Caroline Allenhof https://orcid.org/0000-0001-8329-147X
C. Vis https://orcid.org/0000-0001-8783-8487

Supplemental Material

Supplemental material for this article is available online.

References

Freund, J., Piotrowski, A., Buhrmann, L., Oehler, C., Titzler, I., Netter, A. L., Potthoff, S., Ebert, D. D., Finch, T.,

Koberlein-Neu, J., & Etzelmuller, A. (2023). Validation of the German Normalisation Process Theory Measure G-NoMAD: translation, adaptation, and pilot testing. *Implement Sci Commun*, 4(1), 126. https://doi.org/10.1186/s43058-023-00505-4

Baker, R., Camosso Stefinovic, J., Gillies, C., Shaw, E. J., Cheater, F., Flottorp, S. A., Robertson, N., Wensing, M., Fiander, M., Eccles, M. P., Godycki-Cwirko, M., van Lieshout, J., & Jäger, C. (2015). *Tailored interventions to address determinants of practice* (Vol. 4). John Wiley & Sons, Ltd. http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD005470.pub3/full

Bomfim, R. A., Braff, E. C., & Frazao, P. (2020). Cross-cultural adaptation and psychometric properties of the Brazilian–Portuguese version of the organizational readiness for implementing change questionnaire. *Revista Brasileira De Epidemiologia = Brazilian Journal of Epidemiology*, 23, e200100. https://doi.org/10.1590/1980-549720200100 (Adaptacao transcultural e propriedades psicometricas da versao em portugues (Brasil) do questionario Prontidao Organizacional para Implementacao de Mudanca para implementacao de mudanca em servicos de saude.)

Buhrmann, L., Schuurmans, J., Ruwaard, J., Fleuren, M., Etzelmuller, A., Piera-Jimenez, J., Finch, T., Rapley, T., Potthoff, S., Aouizerate, B., Batterham, P. J., Calear, A., Christensen, H., Pedersen, C. D., Ebert, D. D., Van Der Eycken, E., Fanaj, N., Van Genugten, C., & Hanssen, D., ... ImpleMentAll c. (2020). Tailored implementation of internet-based cognitive behavioural therapy in the multinational context of the ImpleMentAll project: a study protocol for a stepped wedge cluster randomized trial. *Trials*, 21(1), 893. https://doi.org/10.1186/s13063-020-04686-4

Damschroder, L. J., Reardon, C. M., Widerquist, M. A. O., & Lowery, J. (2022). The updated consolidated framework for implementation research based on user feedback. *Implementation Science*, 17(1), 75. https://doi.org/10.1186/s13012-022-01245-0

Elf, M., Nordmark, S., Lyhagen, J., Lindberg, I., Finch, T., & Aberg, A. C. (2018). The Swedish version of the normalization process theory measure S-NoMAD: Translation, adaptation, and pilot testing. *Implementation Science*, 13(1), 146. https://doi.org/10.1186/s13012-018-0835-5

Fernandez, M. E., Ten Hoor, G. A., van Lieshout, S., Rodriguez, S. A., Beidas, R. S., Parcel, G., Ruiter, R. A. C., Markham, C. M., & Kok, G. (2019). Implementation mapping: Using intervention mapping to develop implementation strategies. *Frontiers in Public Health*, 7, 158. https://doi.org/10.3389/fpubh.2019.00158

Finch, T. L., Rapley, T., Girling, M., Mair, F. S., Murray, E., Treweek, S., Mccoll, E., Steen, I. N., & May, C. R. (2013). Improving the normalization of complex interventions: measure development based on normalization process theory (NoMAD): Study protocol. *Implementation science*, 8(1), 43. https://doi.org/10.1186/1748-5908-8-43

Finch, T. L., Girling, M., May, C. R., Mair, F. S., Murray, E., Treweek, S., McColl, E., Steen, I. N., Cook, C., Vernazza, C. R., Mackintosh, N., Sharma, S., Barbery, G., Steele, J., & Rapley, T. (2018). Improving the normalization of complex interventions: part 2 - validation of the NoMAD instrument for assessing implementation work based on normalization process theory (NPT). BMC Med Res Methodol, 18(1), 135. https://doi.org/10.1186/s12874-018-0591-x

Fishbein, M., & Ajzen, I. (1977). *Belief, attitude, intention, and behavior: An introduction to theory and research.*

Gist, M. E., & Mitchell, T. R. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. *Academy of Management Review*, *17*(2), 183–211. https://doi.org/10.2307/258770

Glasgow, R. E. (2013). What does it mean to be pragmatic? Pragmatic methods, measures, and models to facilitate research translation. *Health Education & Behavior*, *40*(3), 257–265. https://doi.org/10.1177/1090198113486805

Glasgow, R. E., Green, L. W., Klesges, L. M., Abrams, D. B., Fisher, E. B., Goldstein, M. G., Hayman, L. L., Ockene, J. K., & Orleans, C. T. (2006). External validity: We need to do more. *Annals of Behavioral Medicine*, *31*(2), 105–108. https://doi.org/10.1207/s15324796abm3102_1

Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health*, *89*(9), 1322–1327. https://doi.org/papers3://publication/doi/10.2105/AJPH.89.9.1322

Grimshaw, J. M., Thomas, R. E., MacLennan, G., Fraser, C., Ramsay, C. R., Vale, L., Whitty, P., Eccles, M. P., Matowe, L., Shirran, L., Wensing, M., Dijkstra, R., & Donaldson, C. (2004). Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technology Assessment*, *8*(6), iii–iv, 1–72. https://doi.org/papers3://publication/uuid/CC192253-E816-43D3-828C-77F227D9DBC1

Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424. https://doi.org/10.1037/1082-989X.3.4.424

Jiang, M., Wang, Q., Finch, T., She, D., Zhou, Y., Chung, Y. F., Chen, J., & Han, L. (2022). Validity and reliability of the Chinese version of the Normalization MeAsure Development(NoMAD). *BMC Health Services Research*, *22*(1), 1338. https://doi.org/10.1186/s12913-022-08737-1

Krause, J., Van Lieshout, J., Klomp, R., Huntink, E., Aakhus, E., Flottorp, S., Jaeger, C., Steinhaeuser, J., Godycki-Cwirko, M., Kowalczyk, A., Agarwal, S., Wensing, M., & Baker, R. (2014). Identifying determinants of care for tailoring implementation in chronic diseases: An evaluation of different methods. *Implementation Science*, *9*(1), 102. https://doi.org/10.1186/s13012-014-0102-3

Lewis, C. C., Fischer, S., Weiner, B. J., Stanick, C., Kim, M., & Martinez, R. G. (2015a). Outcomes for implementation science: An enhanced systematic review of instruments using evidence-based rating criteria. *Implementation Science*, *10*(1), 155. https://doi.org/10.1186/s13012-015-0342-x

Lewis, C. C., Stanick, C. F., Martinez, R. G., Weiner, B. J., Kim, M., Barwick, M., & Comtois, K. A. (2015b). The society for implementation research collaboration instrument review project: A methodology to promote rigorous evaluation. *Implementation Science*, *10*(1), 2. https://doi.org/10.1186/s13012-014-0193-x

Lindig, A., Hahlweg, P., Christalle, E., & Scholl, I. (2020). Translation and psychometric evaluation of the German version of the organisational readiness for implementing change measure (ORIC): A cross-sectional study. *BMJ Open*, *10*(6), e034380. https://doi.org/10.1136/bmjopen-2019-034380

Loch, A. P., Finch, T., Fonsi, M., & Soarez, P. C. (2020). Cross-cultural adaptation of the NoMAD questionnaire to Brazilian Portuguese. *Revista da Associacao Medica Brasileira (1992)*, *66*(10), 1383–1390. https://doi.org/10.1590/1806-9282.66.10.1383

Mansolf, M., & Reise, S. P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence*, *61*, 120–129. https://doi.org/10.1016/j.intell.2017.01.012

Martinez, R. G., Lewis, C. C., & Weiner, B. J. (2014). Instrumentation issues in implementation science. *Implementation Science*, *9*(1), 118. https://doi.org/10.1186/s13012-014-0118-8

May, C. R., Cummings, A., Girling, M., Bracher, M., Mair, F. S., May, C. M., Murray, E., Myall, M., Rapley, T., & Finch, T. (2018). Using normalization process theory in feasibility studies and process evaluations of complex healthcare interventions: A systematic review. *Implementation Science*, *13*(1), 80. https://doi.org/10.1186/s13012-018-0758-1

May, C. R., Mair, F., Finch, T., MacFarlane, A., Dowrick, C., Treweek, S., Rapley, T., Ballini, L., Ong, B. N., Rogers, A., Murray, E., Elwyn, G., Legare, F., Gunn, J., & Montori, V. M. (2009). Development of a theory of implementation and integration: Normalization process theory. *Implementation Science*, *4*, 29. https://doi.org/10.1186/1748-5908-4-29

McHugh, S. M., Riordan, F., Curran, G. M., Lewis, C. C., Wolfenden, L., Presseau, J., Lengnick-Hall, R., & Powell, B. J. (2022). Conceptual tensions and practical trade-offs in tailoring implementation interventions. *Frontiers in Health Services*, *2*. https://doi.org/10.3389/frhs.2022.974095

Mettert, K., Lewis, C., Dorsey, C., Halko, H., & Weiner, B. (2020). Measuring implementation outcomes: An updated systematic review of measures' psychometric properties. *Implementation Research and Practice*, *1*. https://doi.org/10.1177/2633489520936644

Meyer, J. P., & Herscovitch, L. (2001). Commitment in the workplace: Toward a general model. *Human Resource Management Review*, *11*(3), 299–326. https://doi.org/10.1016/S1053-4822(00)00053-X

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*(1), 115–132. https://doi.org/10.1007/BF02294210

Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implementation Science*, *10*, 53. https://doi.org/10.1186/s13012-015-0242-0

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health*, *38*(2), 65–76. https://doi.org/10.1007/s10488-010-0319-7

Proctor, E. K., Powell, B. J., & McMillen, J. C. (2013). Implementation strategies: Recommendations for specifying and reporting. *Implementation Science*, *8*(1), 139. https://doi.org/10.1186/1748-5908-8-139

Rapley, T., Girling, M., Mair, F. S., Murray, E., Treweek, S., McColl, E., Steen, I. N., May, C. R., & Finch, T. L. (2018). Improving the normalization of complex interventions: part 1 - development of the NoMAD instrument for assessing implementation work based on normalization process theory (NPT). BMC Med Res Methodol, 18(1), 133. https://doi.org/10.1186/s12874-018-0590-y

Ruest, M., Léonard, G., Thomas, A., Desrosiers, J., & Guay, M. (2019). French cross-cultural adaptation of the organizational readiness for implementing change (ORIC). *BMC Health Services Research*, *19*, 1–14. https://doi.org/10.1186/s12913-019-4361-1

Shea, C. M., Jacobs, S. R., Esserman, D. A., Bruce, K., & Weiner, B. J. (2014). Organizational readiness for implementing change: A psychometric assessment of a new measure. *Implementation Science*, *9*(1), 7. https://doi.org/10.1186/1748-5908-9-7

Storkholm, M. H., Mazzocato, P., Tessma, M. K., & Savage, C. (2018). Assessing the reliability and validity of the danish version of organizational readiness for implementing change (ORIC). *Implementation Science*, *13*(1), 78. https://doi.org/10.1186/s13012-018-0769-y

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Vis, C., Schuurmans, J., Aouizerate, B., Craggs, M., Batterham, P., Buhrmann, L., Calear, A., Cerga Pashoja, A., Christensen, H., Dozeman, E., Pedersen, C., Ebert, D. D., Etzelmueller, A., Fanaj, N., Finch, T. L., Hanssen, D., Hegerl, U., Hoogendoorn, A., Mathiasen, K., & , … Riper, H. (2023). Effectiveness of Self-guided Tailored Implementation Strategies in Integrating and Embedding Internet-Based Cognitive Behavioral Therapy in Routine Mental Health Care: Results of a Multicenter Stepped-Wedge Cluster Randomized Trial. *Journal of Medical Internet Research*, *25*, e41532. https://doi.org/10.2196/41532

Vis, C., Mol, M., Kleiboer, A., Buhrmann, L., Finch, T., Smit, J., & Riper, H. (2018). Improving Implementation of eMental Health for Mood Disorders in Routine Practice: Systematic Review of Barriers and Facilitating Factors. JMIR Ment Health, *5*(1), e20. https://doi.org/10.2196/mental.9769

Vis, P. D. C., Ruwaard, J., Finch, T., Rapley, T., de Beurs, D., van Stel, H., van Lettow, B., Mol, M., Kleiboer, A., Riper, H., & Smit, J. H. (2019). Toward an Objective Assessment of Implementation Processes for Innovations in Health Care: Psychometric Evaluation of the Normalization Measure Development (NoMAD) Questionnaire Among Mental Health Care Professionals. J Med Internet Res, *21*(2), e12376. https://doi.org/10.2196/12376

Vroom, V. H. (1964). *Work and motivation*. John Wiley & Sons, Inc.

Weiner, B. J. (2009). A theory of organizational readiness for change. *Implementation Science*, *4*(1), 67. https://doi.org/10.1186/1748-5908-4-67

Weiner, B. J., Clary, A. S., Klaman, S. L., Turner, K., & Alishahi-Tabriz, A. (2020a). Organizational readiness for change: What we know, what we think we know, and what we need to know. In *Implementation Science 3.0* (pp. 101–144). https://doi.org/10.1007/978-3-030-03874-8_5

Weiner, B. J., Lewis, C. C., Stanick, C., Powell, B. J., Dorsey, C. N., Clary, A. S., Boynton, M. H., & Halko, H. (2017). Psychometric assessment of three newly developed implementation outcome measures. *Implementation Science*, *12*(1), 108. https://doi.org/10.1186/s13012-017-0635-3

Weiner, B. J., Mettert, K. D., Dorsey, C. N., Nolen, E. A., Stanick, C., Powell, B. J., & Lewis, C. C. (2020b). Measuring readiness for implementation: A systematic review of measures' psychometric and pragmatic properties. *Implement Res Pract*, *1*, 2633489520933896. https://doi.org/10.1177/2633489520933896

Wensing, M. (2017). The tailored implementation in chronic diseases (TICD) project: Introduction and main findings. *Implementation Science*, *12*(1), 5. https://doi.org/10.1186/s13012-016-0536-x