PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Translation and psychometric evaluation of the German version of the Organizational Readiness for Implementing Change measure (ORIC) – a cross-sectional study
AUTHORS	Lindig, Anja; Hahlweg, Pola; Christalle, Eva; Scholl, Isabelle

VERSION 1 – REVIEW

REVIEWER	Melanie Ruest Université de Sherbrooke, Québec (Canada)
REVIEW RETURNED	05-Nov-2019

GENERAL COMMENTS	Thank you for the opportunity to review this paper. I have chosen to provide major points followed by free-text comments for each section of the paper, with the reference to specific lines when it was relevant.
	Short summary
	The purpose of the manuscript is to present the translation, adaptation and psychometric testing of a German version of the Organizational Readiness for Implementing Change (ORIC). Based on the TRAPD (Translation, Review, Adjudication, Pretesting and Documentation) translation protocol and secondary analyses of data collected in a previous implementation study, this paper details the translation, comprehensibility assessment as well as psychometric evaluations (e.g., structural and content validity, reliability) of the new German version of the ORIC.
	Overall advice
	In a context of implementation studies that increasingly recognized the relevance of measuring the organizational readiness to help identifying the next steps to take for the facilitation process, this new German version of a well-documented questionnaire (i.e., ORIC) has the potential to help German organizations in their implementation efforts. Firstly, I want to acknowledge the quality of this manuscript and congratulate the authors for this work. The process presented in this manuscript testify to the rigor with which the authors want to share their research. This paper has therefore

a great potential for publication. However, some concerns should be considered in my view before the publication of this paper, as detailed below.

Major points

Abstract

Methodology

Line 34: In order to better represent the content of this paper, the authors are invited to review the identification of the design.

- The translation process is usually considered as a methodological design in itself when using a specific protocol.
- A psychometric study seems however not well-suited to be presented as a design. For example, would the SDM implementation study being considered as an observational study?

Line 44: Since only the internal consistency has been calculated in this manuscript and reliability refers to many different aspects (e.g., interrater agreement, test-retest reliability), I would suggest to the authors to specifically refer to the internal consistency.

Major points

Manuscript

According to the vocabulary used by Weiner and Shea, the second aspect of organizational readiness tends to be defined as change *efficacy*. Throughout the manuscript, I noticed that the authors instead used the concept "change *efficiency*". They are invited to justify this difference or correct the word choice if necessary (e.g., lines 95, 114 and 172).

Objectives:

In the abstract, the authors identified three objectives. Among them, the objective of adapting the ORIC to the context of shared decision-making (SDM) is presented following the translation process. On the contrary, in light of the reading of the methodology and results described, I rather understand that these adaptations were made simultaneously with the translation process. Even though the results of the cognitive interviews supported effectively the adaptation of the scale, their use seems to be mainly in line with the understanding of the German version of the ORIC rather than specifically for the SDM context.

To guide the reflection about the potential of clarifying or reordering some aspects of the translation and adaptation processes described in the manuscript, I will raise two points:

> (1) If the adaptation of the questionnaire to the SDM implementation study still a relevant objective to present according to the authors, the methodology described to this end should be presented clearlier, since the documentation of content validity constitutes a part of the psychometric studies usually performed.

> If not, I would suggest to the authors to reformulate the objectives around the translation process and the documentation of its psychometric qualities, the two main themes of this paper in my view.

> (2) Could the authors discuss the challenges related to having performed a translation process of a tool already adapted to the SDM implementation study for the potential of generalizability of the produced German ORIC version for other potential organizations or studies? Some anticipated limitations should be presented in the Discussion section.

Methods and Data analysis:

As the psychometric evaluation well-described, the translation process performed could be further detailed (lines 121-126). For example, could the authors detail the steps of the translation protocol TRAPD (Translation, Review, Adjudication, Pretesting and Documentation)? Has a back translation process been performed before the final review of the German version by the research team? What is the experience of the translators involved in the process? Who participated to the discussions for developing the final version? This information would be important to present in

order to facilitate the understanding of the translation process performed.

The original English version of the ORIC uses a 5-point Likert scale ranging from 1 (Disagree) to 5 (Agree). The authors are invited to justify their choice of changing the rating system of the questionnaire, as stated at line 113.

Concerning the psychometric analyses, my main concern is about the choice of performing an orthogonal rotation for the exploratory factor analysis. Since the *Commitment* and *Efficacy* subscales of the original version of the ORIC, based on Weiner's theory, are conceptualized as being correlated to one point or another to define the organizational readiness, the postulate of independence for this type of rotation is not respected. The authors have to justify their choice of performing such analyses.

Results and Discussion:

In line with the previous point about the description of translation process in the Methods section, the authors are invited to detail the results of the translation process and discuss moreover them. For example, how many translation differences have been identified? How have they been handled?

In the Results section, the authors refer to several tables regarding the confirmatory and exploratory factor analyses performed. However, several of these results are not discussed in this section (e.g., corrected item-total correlation, inter-item correlations). Even if not all the results need to be addressed, the authors are invited to discuss at least some of them in light of the criteria presented in Table 1.

The psychometric measures tend to present arguments both for the use of the 9 and 10-item versions of the ORIC. The argumentation presented in the discussion regarding the decision to promote the use of a 9-item German ORIC version could be further detailed and triangulated regarding the comprehensibility and psychometric measures used since as a reader, I am not convinced about the version of the German questionnaire that I should use. The authors are therefore invited to deepen their reasoning about their recommendation to use the 9-item version of the ORIC.

Finally, in line with the suggestion of clarifying the idea of having translate the ORIC questionnaire already modified for the SDM implementation context of study (see Objectives section above), the authors are invited to discuss the implications of it (e.g., in terms of generalizability for future contexts of use in German organizations).

Minor points

I made a throughout reading of the manuscript, but since English is my second language, I am not qualified to assess the quality of the language in this paper. However, I invite the authors to perform another review for appropriately-worded English as some sentences' structures were not optimal to facilitate the understanding of the reader (e.g., absent or unnecessary [line 117] commas, term *adaptation* instead of *adaption*, verb tenses [line 288], and grammatical chords [e.g., line 129]).

Minor points

Abstract

Objectives

Line 33: The abbreviation "SDM" should be detailed since it is the first time it is used in the manuscript.

Conclusions

Line 51: Please review the structure of the following excerpt: "[…] with satisfying psychometric properties **and a one factorial structure**".

Minor points

Manuscript

Article summary

Lines 68-69: The authors are invited to give examples of psychometric parameters that could not be calculated.

Introduction

Suggestion: The multiple paragraphs could be grouped to facilitate the contextualization of this section (e.g., paragraphs of lines 74-77 and 78-83 and paragraphs of lines 84-88 and 89-97).

Lines 81-83: The shared decision-making (SDM) could be further detailed in order to support its relevance. For example, the authors could refer to specific results of researches or health policies advocating for its use.

Line 95: Please review the structure of the following excerpt to follow the nomenclature used with the questionnaire. Here a suggestion: "[…] and change efficiency, where they belief in capability to **implementing a** change".

Line 97: Which difference(s) do the authors make between obstacles and setbacks? The consistency on the vocabulary used throughout the manuscript facilitates understanding.

Lines 101-102: The authors are invited to relativize the following sentence: "The ORIC is the only available measure that is brief, easy to administer and theoretically grounded."

For instance, there are other questionnaires in the field of organizational readiness with these characteristics, but the ORIC is particularly well-grounded from theoretical and psychometric perspectives.

Methods

Lines 137-138: The specification of the compensation fee appears in our opinion unnecessary information to present for the understanding of readers or justify the scientifically robustness. Please omit.

Some information are also duplicated in the Methods and Results sections. For example, the involvement of a second bilingual

researcher (DF) presented at lines 225-226 is already detailed at line 143. Do the authors refer to two distinct steps in the translation process?

Line 138: Could the authors give an example of questions used in the cognitive interview guide or an example of Willis' recommendations considered for the development of their questionnaire?

Line 142: The authors specified at this point in the manuscript that the items would be discussed by the study team. Since the items are not specified and no details are presented to this end in the Results section, the information provided is partial. Please detail the items discussed and the conclusions that emerged from it between the members of the research team.

Lines 147-148: The information about the descriptive statistics should be presented in the Data Analysis section.

Line 156: The authors are invited to reformulate the following excerpt: "[…] as well as several self-generated questions."

Suggestion: "[…] as well as several questions based on a previous oncologic study".

Line 158: Were the psychooncologists not supposed to be also participants, as stated at lines 135-136? Please clarify.

Lines 160-163: The recruitment process appears unclear. Is this a convenient sampling where the members of the study team recruit participants while being patients during their physician meetings?

Data analysis

Lines 166-167: The authors are invited to specify the underlying rational for establishing the % cut-off of missing items to decide the exclusion (or not) of cases and why they would include them as well as for the analysis of acceptance.

Lines 200-205: The paragraph "Patient and public involvement" is written with a different police style. Also, since this information is

about the recruitment process of the study, I suggested to the authors to present it previously (e.g., after line 163).

Results

Line 218: Has the duration of cognitive interviews been calculated? If so, the authors could present the mean duration and the standard deviation.

Line 242: The authors are invited to standardize the unit to which the percentages are presented in the Table 2 (e.g., % of age [<30 years] and % of gender [different gender or preferred not to answer questions]).

Lines 269-270: The authors report missing data for several items of the questionnaire. Could the authors discuss about the implications related to the data collection procedures used?

Line 274: Same comment than the one made for line 44: I would suggest to the authors to specifically refer to the internal consistency considering this is the only aspect of reliability that has been studied.

Line 279: At the end of the Table 5, the authors specified: "For calculation of response rate, four additional cases were included because these participants only skipped the ORIC but filled out the rest of the questionnaire."

This specification could be relevant for future publications with the data collected in the SDM implementation study. However, for this present paper focusing specifically on the ORIC, I do not understand the relevance of including such data.

Discussion

Lines 281-285: Even though the translation process constitutes the first objective of the paper and has been detailed in the Methods section, no paragraph about this process is presented at the beginning of this present section. As stated previously, the authors are therefore invited to detail the results of the translation process and triangulated them (strengths/limitations).

References

Line 437: The volume of the reference is missing.

Line 448: The volume of the reference is double (123-124) and there is a letter at the end (A).

Line 458: The volume of the reference is double (123-124).

Lines 464-465: The title of this reference is the only one presented in brackets.

Line 479: The place of publication of this reference is missing.

Line 521: The page numbers of this reference are missing.

Lines 527 and 538-539: Some information for these references are missing (publication house and place).

Line 547: The volume and the page numbers of the reference are missing.

REVIEWER	Mesfin Kassaye Tessma Karolinska Institutet, Sweden
REVIEW RETURNED	02-Dec-2019

| GENERAL COMMENTS | **Abstract:**

Objective: Please check the objectives stated in the text (page 4, line 107). It is not consistent.

Design: It is unclear and not specific. Please use a better description (secondary analysis of …)
Primary and Secondary Outcome Measures: This part included data analysis (not consistent with the subtitle). Please modify the subtitle or the content. SDM (?) – please specify.
Results: CFA is mentioned as a statistical method in the abstract but in the "Results" part of the abstract results related to CFA were not given. Results from EFA do not provide evidence for construct validity. I suggest to include results from the CFA (if "the German ORIC can be used to analyze organizational readiness for change" |

– as the authors recommended in the conclusion part). The finding "low comprehensibility" is unclear. In the text it is defined but difficult to understand in the abstract. This applies also to structural validity. Item difficulty is often used in Rasch analysis not in factor analysis. It can be confusing. It seems as if the authors have performed Rasch analysis.

Conclusion: The phrase "highly accepted" may be misleading.

Introduction

The use of the terms "good reliability" and "good validity" (page 4, lines 104-105) and the "values" related to the definitions of these terms are often confusing. I don't think the references indicated support the use of these terms and their "acceptable values". Please modify.

Methods

Subtitle "Psychometric evaluation" (page 9, line 234): "psychometric evaluation" is a better description (in the title for example, it is "psychometric testing")

Data analysis

Missed values (page 6, line 167): The authors have stated that cases were excluded if more than 30% of the ORIC items were missing. Why 30%? They have also replaced the item mean for all missed values (but often multiple imputation is used). Why not multiple imputation instead of mean? The issue of missing values is not as such a problem, if Rasch analysis was used. See my comments under "General".

Page 6, line 175: "Because the two-factor model could not be confirmed, the data was randomized and split into two subsets (each n =115)". This is very confusing (please see the results of the CFA in Table 4). What does it mean in the light of Table 4? What does it mean that the data set was randomized? How was the randomization process performed? Who performed it? What was the purpose of the randomization?

Page 7, line 190: "Because of low content validity of item 10 …" – this is unclear. Please

The authors have used both EFA and CFA, which is the right approach when using CTT. However, EFA helps in evaluating the dimensionality of scales but it does not provide evidence for construct validity. It is CFR that provides evidence for construct validity. As we know construct validity is the extent to which an instrument measures the construct that it is intended to measure.

Please provide information on how the sample size was determined?

The authors did not use parallel analysis to determine the number of factors to be retained in the EFA. Why?

Results

Table 2: Percentage are given as one or two decimal places. Please round it to one decimal place. Do the data meet the appropriate assumptions for EFA and CFA?

Table 3: Please expand the title of the table – the title identifies the data displayed in the table and the context in which they should be interpreted. Many horizontal lines are used in the table. Do we need them? Why was half of the dataset used for EFA? (Please see footnote)

How does missing values and outliers affect the result? Please comment!

Based on AIC and PNFI results the authors preferred the one factor model, although the two factor model has acceptable and better values for TLI and RMSEA. However, since CFA is a confirmatory technique and it is theory driven, it is very important to relate it to the theoretical model. Therefore, the analysis considers the theoretical relationships among the observed and unobserved variables (Schrebier et al, 2006). Thus it seems the two factor model may be better in the light of the theory. Construct validity is achieved when the measure reflects the framework hypothesized in a hypothesis testing study, adequately establishes that the measurement model fits the actual data. Have the authors considered the theory?

The results of the CFA cannot be presented as sufficient evidence unless the assumptions are checked and the validity of the findings "confirmed". Therefore, to support the findings please check and report about the statistical assumptions of CFA. Please provide information about the assumptions of CFA.

Discussion

Page 13, line 303: Structural validity "we could not confirm the two-factor structure of the English ..." But the findings indicated the absolute and comparative indices and RMSEA indicated that the two factor model is acceptable. Why did the authors reported that they "could not confirm" the two factor model? EFA explores the factor structure and CFA confirms the factor structure extracted in EFA.

Page 13, line 308: "… one factor model **should be** preferred". It is a very strong statement considering the findings. Please modify (please see also my comments above)

REVIEWER	Kira Hower
	University of Cologne, Germany
REVIEW RETURNED	18-Jan-2020

GENERAL COMMENTS	**Review**
	Journal: BMJ Open
	Manuscript ID bmjopen-2019-034380
	I had the opportunity to review the manuscript entitled "Translation, adaption and psychometric testing of the German version of the Organizational Readiness for Implementing Change measure (ORIC)". This study took advantage of a secondary analysis of data gathered in a SDM implementation study in a German hospital to assess the psychometric properties of the German-language measure "ORIC".
	The authors reported that an abbreviated measure reflected the conceptual structure of the construct, convergent validity, as well as item characteristics and reliability. The study has several strengths including that the analytic approach was well-described and appropriate, with an interesting qualitative approach to explore comprehensibility. Unfortunately, the study has some limitations including the cross-sectional nature of the study design, no testing of the criterion-based validity, the fact that the instrument is only applied in one organizational setting and shared method variance as well as the likelihood of halo effect.
	Introduction
	The introduction is well structured. However, a few questions have occurred regarding implementation process, status and outcomes. Has SDM already been implemented in the hospital? From the description of the instrument it can be concluded that there are variations, i.e. SDM may or may not be implemented and known. This should be explained in more detail and further information on the implementation should be provided. If SDM is already implemented, there is a high likelihood of halo effect from this design, i.e., respondents recalling the initiative of SDM implementation and inferring responses to their attitudes to the change based on the outcome. This is a huge problem for retrospective studies that ask respondents to evaluate factors contributing to an outcome that is known.
	The wording "implementation of change" can be confusing, since it is not change that is implemented, but a new practice, structure, etc., that requires changes at different levels. Which changes are required by SDM in hospitals could be explained in more detail. This would make it easier to understand how organizational readiness is affected.

If SDM is already implemented, is there any information on the implementation status that could be related to organizational readiness for change?

Methods

The methodological and statistical procedures were explained in a clear and understandable way and are in line with the standards. Linking the data with other data, e.g. on the status of implementation or other external criteria, would have significantly increased the validity of the findings and would have contributed a clear added value to the research.

A few remaining clarifications should be made.

Have you tested for differences between groups/personal characteristics/departments? Is there evidence in the literature that there are differences?

Do the participants hold leading positions? Has this been taken into account in the survey and analysis? Organizational readiness is clearly influenced by the role in the organization, so this would be, if possible, interesting information to consider.

For how many cases were missing values replaced?

Did the authors make model modifications, such as the correlation of error terms?

Results

The presentation of results is detailed, well structured, and clearly illustrated. Regarding the results themselves I have the following minor concerns that may need to be addressed.

How do the authors explain the fact that in this study, in contrast to the other validation studies, a single factor structure of the scale emerges? What does this mean in terms of a limitation?

Also, knowledge about SDM and the organization must be available to answer the questions. Since the descriptive statistics show that the average values are rather in the middle, it might be that the participants did not know enough to answer the questions? How could this be ensured and would this be an important prerequisite in terms of using the scale?

Discussion

The discussion is well structured, but is brief and could benefit from more explanations and implications. The limitations could be extended, e.g. with regard to the points mentioned above.

In relation to the conclusion, the authors could make it clearer what the theoretical and practical contribution of the study is and what the results contribute to research.

	In conclusion, I consider the manuscript to be worth publishing and recommend some adaptations or additions.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Melanie Ruest
Institution and Country: Université de Sherbrooke, Québec (Canada)
Please state any competing interests or state 'None declared': No competing interests to declare. Please note that one of my published studies is used in the references of this study. I also exchanged emails with an author of this manuscript to discuss about our respective experiences of translation for an item of the questionnaire.

November 5, 2019

Dr Emma Cookson Senior Assistant Editor, BMJ Open BMJ, BMA House, Tavistock Square, London, WC1H 9JR ecookson@bmj.com

Re: Manuscript ID bmjopen-2019-034380 revision Translation, adaption and psychometric testing of the German version of the Organizational Readiness for Implementing Change measure (ORIC)

Dr Cookson,

Thank you for the opportunity to review this paper. I have chosen to provide major points followed by free-text comments for each section of the paper, with the reference to specific lines when it was relevant.

Short summary

The purpose of the manuscript is to present the translation, adaptation and psychometric testing of a German version of the Organizational Readiness for Implementing Change (ORIC). Based on the TRAPD (Translation, Review, Adjudication, Pretesting and Documentation) translation protocol and secondary analyses of data collected in a previous implementation study, this paper details the translation, comprehensibility assessment as well as psychometric evaluations (e.g., structural and content validity, reliability) of the new German version of the ORIC.

Overall advice

In a context of implementation studies that increasingly recognized the relevance of measuring the organizational readiness to help identifying the next steps to take for the facilitation process, this new German version of a well-documented questionnaire (i.e., ORIC) has the potential to help German organizations in their implementation efforts. Firstly, I want to acknowledge the quality of this manuscript and congratulate the authors for this work. The process presented in this manuscript testify to the rigor with which the authors want to share their research. This paper has therefore a great potential for publication. However, some concerns should be considered in my view before the publication of this paper, as detailed below.

Thank you very much for your positive feedback and your suggestions and comments.

<u>Major points Abstract</u>

1. Methodology Line 34: In order to better represent the content of this paper, the authors are invited to review the identification of the design. The translation process is usually considered as a methodological design in itself when using a specific protocol. A psychometric study seems however not well-suited to be presented as a design. For example, would the SDM implementation study being considered as an observational study?
Our study was a cross-sectional study. We added the design to the title, the revised abstract and at the beginning of the revised methods section for the psychometric analysis.

2. Line 44: Since only the internal consistency has been calculated in this manuscript and reliability refers to many different aspects (e.g., interrater agreement, test-retest reliability), I would suggest to the authors to specifically refer to the internal consistency.
Thank you for the suggestion, we changed the term reliability to internal consistency in the whole revised manuscript.

Major points Manuscript

3. According to the vocabulary used by Weiner and Shea, the second aspect of organizational readiness tends to be defined as change efficacy. Throughout the manuscript, I noticed that the authors instead used the concept "change efficiency". They are invited to justify this difference or correct the word choice if necessary (e.g., lines 95, 114 and 172).
This was a mistake; in the revised manuscript we changed it to "efficacy".

Objectives:

4. In the abstract, the authors identified three objectives. Among them, the objective of adapting the ORIC to the context of shared decision-making (SDM) is presented following the translation process. On the contrary, in light of the reading of the methodology and results described, I rather understand that these adaptations were made simultaneously with the translation process. Even though the results of the cognitive interviews supported effectively the adaptation of the scale, their use seems to be mainly in line with the understanding of the German version of the ORIC rather than specifically for the SDM context. –
You are right, the major focus was on the translation process. According to Shea et al. the measure can be specified to the research context, so we adapted to shared decision-making during the translation process and before cognitive interviews. Cognitive interviews did not focus much on the SDM context. We tried to make it clearer in the revised methods section (part "Translation"): "During the translation process, we changed the phrases "to implement this change" and "implementing this change" into "to implement shared decision-making" and "implementing shared decision-making" to address our specific research question. Additionally, we added a survey instruction in German."

5. To guide the reflection about the potential of clarifying or reordering some aspects of the translation and adaptation processes described in the manuscript, I will raise two points: (1) If the adaptation of the questionnaire to the SDM implementation study still a relevant objective to present according to the authors, the methodology described to this end should be presented clearlier, since the documentation of content validity constitutes a part of the psychometric studies usually performed. If not, I would suggest to the authors to reformulate the objectives around the translation process and the documentation of its psychometric qualities, the two main themes of this paper in my view.
We decided to reformulate the objectives in the revised abstract, the revised introduction and the revised discussion by leaving out the objective "to adapt it for the context of SDM". We also changed the title to: "Translation and psychometric evaluation of the German version of the Organizational Readiness for Implementing Change measure (ORIC) – a cross-sectional study." The focus of this study is more on translation and psychometric evaluation, the adaptation for the context of SDM is more part of the translation process and therefore not a central objective (see comment 4).

6. (2) Could the authors discuss the challenges related to having performed a translation process of a tool already adapted to the SDM implementation study for the potential of generalizability of the produced German ORIC version for other potential organizations or studies? Some anticipated limitations should be presented in the Discussion section.
Thank you for the important suggestions, we added that point to the revised discussion of limitations: "Third, for this psychometric evaluation we used a German ORIC, which we adapted and specified for the context of SDM implementation. Our results might not be generalizable for other interventions in other organizations."

<u>Methods and Data analysis:</u>

7. As the psychometric evaluation well-described, the translation process performed could be further detailed (lines 121-126). For example, could the authors detail the steps of the translation protocol TRAPD (Translation, Review, Adjudication, Pretesting and Documentation)? Has a back translation process been performed before the final review of the German version by the research team? What is the experience of the translators involved in the process? Who participated to the discussions for developing the final version? This information would be important to present in order to facilitate the understanding of the translation process performed.
We added more detailed information about the background of the TRAPD protocol to the revised methods section (part "Translation"): "Translation followed the team translation protocol TRAPD (Translation, Review, Adjudication, Pretesting and Documentation), a method with growing recognition within translation research [48]. Thereby an optimal translation is facilitated by discussion between members of the translation team with different expertise in translation." We did not perform a back translation. We also added details about translator experience in this revised section as well as participants of the discussion of the final version: "IS, AL and SZ discussed all versions until reaching consensus on a final version".

8. The original English version of the ORIC uses a 5-point Likert scale ranging from 1 (Disagree) to 5 (Agree). The authors are invited to justify their choice of changing the rating system of the questionnaire, as stated at line 113.
Thank you for your comment. You are right, we used a scale ranging from 0 to 4. Since this is also a 5-point Likert scale, it is comparable to other ORIC versions using other scale values. When comparing data with data from a study using a scale ranging from 1 to 5, once only have to add +1 for the mean values, other reported values can be directly compared.

9. Concerning the psychometric analyses, my main concern is about the choice of performing an orthogonal rotation for the exploratory factor analysis. Since the Commitment and Efficacy subscales of the original version of the ORIC, based on Weiner's theory, are conceptualized as being correlated to one point or another to define the organizational readiness, the postulate of independence for this type of rotation is not respected. The authors have to justify their choice of performing such analyses.
Thank you very much for this remark. We agree with you. A non-orthogonal rotation would have been more applicable based on the theory. We recalculated the factor analysis with an oblique rotation and parallel analysis, but as we extracted only one factor, this did not change our previous results. We revised the method section: "An EFA with oblique rotation was calculated for the first subset. The non-orthogonal rotation was chosen according to Weiner et al. [40]. In their theory, organizational readiness for change consists of two interrelated dimensions, therefore the two factors are expected to be correlated. Analogue to analysis done by authors of the English ORIC [43,61], we extracted components based on parallel analysis. The criterion of parallel analysis was shown to be superior to other statistic criteria like the Kaiser criterion [62,63]. It compares the eigenvalues of the data to the eigenvalues based on random data with equivalent sample size and number of variables and chooses only factors with eigenvalues higher than for random data [61,62]."

Results and Discussion:

10. In line with the previous point about the description of translation process in the Methods section, the authors are invited to detail the results of the translation process and discuss moreover them. For example, how many translation differences have been identified? How have they been handled?
Thank you for the suggestion, we revised the results section and added a lot more details on the translation process, translation differences and specific adaptations made within both rounds of cognitive interviews.

11. In the Results section, the authors refer to several tables regarding the confirmatory and exploratory factor analyses performed. However, several of these results are not discussed in this section (e.g., corrected item-total correlation, inter-item correlations). Even if not all the results need to be addressed, the authors are invited to discuss at least some of them in light of the criteria presented in Table 1.
We improved that and explored more results in the results section (especially results of psychometric analysis of the 9-item ORIC version): "For the 9-item version also, a one-factor model was assumed by exploratory factor analysis with a main component explaining 66.85% of the variance and factor loadings in the first component of above 0.739 for all items (see Supplementary File 4, Table A and B)." and "Corrected item-total correlations and internal consistency were again calculated for the 9-item ORIC version (see Supplementary File 4, Table D). They have shown to be similar compared to results for the 10-item version with corrected item-total correlations between .638 (item 9) and .777 (item 3) and Cronbach's α of .92.". We also discussed results of item analysis in more detail in the discussion section: "There were no floor or ceiling effects. Corrected item-total correlations of above .66 indicate that all items measure the same underlying concept [57,69]. Criteria for good item difficulties are met since item difficulties are near to 50% and do not differ much from each other [58]. Inter-item correlations are below .80, indicating that items add additional information and are not redundant [57,69]. Cronbach's α (α=.93) suggest excellent internal consistency [57,58,70]. In summary, item analysis and internal consistency of the German ORIC suggest good quality of the measure."

12. The psychometric measures tend to present arguments both for the use of the 9 and 10-item versions of the ORIC. The argumentation presented in the discussion regarding the decision to promote the use of a 9-item German ORIC version could be further detailed and triangulated regarding the comprehensibility and psychometric measures used since as a reader, I am not convinced about the version of the German questionnaire that I should use. The authors are therefore invited to deepen their reasoning about their recommendation to use the 9-item version of the ORIC.
Indeed, the results on content validity led to our suggestion to leave out item 10, also because the 9-item version has very similar psychometric properties than the 10-item version. But since we did not test the 9-item version in a separate sample, a "recommendation" to leave the item out might be too strong. We discussed that again in the research team and made some adaptions to the discussion section: "To increase comprehensibility and therefore content validity of the scale, the use of the 9-item German ORIC might be a solution and should be evaluated in future studies."

13. Finally, in line with the suggestion of clarifying the idea of having translate the ORIC questionnaire already modified for the SDM implementation context of study (see Objectives section above), the authors are invited to discuss the implications of it (e.g., in terms of generalizability for future contexts of use in German organizations).

We added this suggestion to the revised discussion section (part "Strengths and Limitations"), please see comment 6.

Minor points

14. I made a throughout reading of the manuscript, but since English is my second language, I am not qualified to assess the quality of the language in this paper. However, I invite the authors to perform another review for appropriately-worded English as some sentences' structures were not optimal to facilitate the understanding of the reader (e.g., absent or unnecessary [line 117] commas, term adaptation instead of adaption, verb tenses [line 288], and grammatical chords [e.g., line 129]).
We carefully proofread the manuscript again.

Minor points Abstract

15. Objectives Line 33: The abbreviation "SDM" should be detailed since it is the first time it is used in the manuscript.
We added "shared decision-making" to the revised abstract when mentioning SDM the first time.

16. Conclusions Line 51: Please review the structure of the following excerpt: "[…] with satisfying psychometric properties and a one factorial structure".
We changed that in the revised abstract: "The German ORIC is a reliable measure with high completition rates and satisfying psychometric properties. A one-factorial structure of the German ORIC was confirmed".

Minor points Manuscript

Article summary

17. Lines 68-69: The authors are invited to give examples of psychometric parameters that could not be calculated.
We added the following examples to the revised manuscript: "convergent and divergent validity".

Introduction

18. Suggestion: The multiple paragraphs could be grouped to facilitate the contextualization of this section (e.g., paragraphs of lines 74-77 and 78-83 and paragraphs of lines 84-88 and 89-97).
We grouped the paragraphs of the revised introduction to three paragraphs.

19. Lines 81-83: The shared decision-making (SDM) could be further detailed in order to support its relevance. For example, the authors could refer to specific results of researches or health policies advocating for its use.
Thank you for suggestion, we added some further details on relevance of SDM in the revised introduction: "This is especially important in situations with complex treatment options and high impact on patients' quality of life [15]. Patients want to be actively involved in decision-making [16] and benefit from SDM by developing better knowledge about their disease and treatment, better risk perception and less insecurity and decisional conflict [17,18]."

20. Line 95: Please review the structure of the following excerpt to follow the nomenclature used with the questionnaire. Here a suggestion: "[…] and change efficiency, where they belief in capability to implementing a change".
We changed the sentence according to your suggestion.

21. Line 97: Which difference(s) do the authors make between obstacles and setbacks? The consistency on the vocabulary used throughout the manuscript facilitates understanding.

We revised the manuscript to improve consistency of vocabulary: "If readiness for change is high, organizational members invest more in the change effort and exhibit greater persistence to overcome barriers and setbacks [40,41]."

22. Lines 101-102: The authors are invited to relativize the following sentence: "The ORIC is the only available measure that is brief, easy to administer and theoretically grounded." For instance, there are other questionnaires in the field of organizational readiness with these characteristics, but the ORIC is particularly well-grounded from theoretical and psychometric perspectives.
Thank you, you are right, we specified this sentence in the revised introduction as follows: "The ORIC is brief, easy to administer and theoretically and psychometrically well-grounded."

<u>Methods</u>

23. Lines 137-138: The specification of the compensation fee appears in our opinion unnecessary information to present for the understanding of readers or justify the scientifically robustness. Please omit.
We omitted this sentence in the revised methods section.

24. Some information are also duplicated in the Methods and Results sections. For example, the involvement of a second bilingual researcher (DF) presented at lines 225-226 is already detailed at line 143. Do the authors refer to two distinct steps in the translation process?
DF was involved once for discussion of item 10 after the second round of cognitive interviews. We tried to write that clearer. We explained the involvement of DF in the methods section (part "Assessment of comprehensibility"): "After the second round, we again discussed further adaptions of the items and therefore involved another bilingual researcher in the field (DF, cp. list of abbreviations)." and we explained results of the discussion with DF in the results section (part "Assessment of comprehensibility"): "Both rounds of cognitive interviews showed that comprehension of the German translation of the phrase "manage the politics" did not picture the correct English meaning. So we involved DF (cp. list of abbreviations) after conducting cognitive interviews, and found consensus for an adapted version of item 10."

25. Line 138: Could the authors give an example of questions used in the cognitive interview guide or an example of Willis' recommendations considered for the development of their questionnaire?
Yes, we specified some examples of the verbal probing technique in the revised methods section (part "Assessment of comprehensibility"): "We developed an interview guide based on recommendations of Willis et al. [51]. We used verbal probing techniques like comprehension probes (e.g. "What does the term "organization" mean to you?") and paraphrasing (e.g. "Can you repeat this sentence in your own words?")."

26. Line 142: The authors specified at this point in the manuscript that the items would be discussed by the study team. Since the items are not specified and no details are presented to this end in the Results section, the information provided is partial. Please detail the items discussed and the conclusions that emerged from it between the members of the research team.
We revised the description of the translation and adaption process in the methods section and the results section (see also our answer on your comment 10). We included results of discussion of the study team members for the translation process (e.g. for which item did we found a fast consent, which items needed several rounds of discussions) and the adaptation process (e.g. which items were understood well by all participants of the first rounds of cognitive interviews, which items had to be adapted by the study team again and had to be tested in a second round of cognitive interviews, for which items did we test several versions).

27. Lines 147-148: The information about the descriptive statistics should be presented in the Data Analysis section.

These are information about the descriptive data analysis of the participants of cognitive interviews. The data analysis section refers only to the participants in the psychometric analysis. Therefore, we would suggest not to shift these information to another section.

28. Line 156: The authors are invited to reformulate the following excerpt: "[…] as well as several self-generated questions." Suggestion: "[…] as well as several questions based on a previous oncologic study".
Thank you, we changed this sentence in the revised methods section to: "Besides the ORIC, it contained the Control Preference Scale [53] and the IcanSDM [54], demographic questions (e.g. gender, age, profession, work experience) as well as several questions that have been used in previous studies in cancer care [55]."

29. Line 158: Were the psychooncologists not supposed to be also participants, as stated at lines 135136? Please clarify.
Thank you for your question. Psychooncologists were invited to participate in the cognitive interviews but were not part of the secondary analysis due to predefined inclusion criteria of the SDM implementation study. We specified this in the revised data collection section: "Participants were part of a convenience sample of physicians and nurses. Since this is a secondary analysis, inclusion criteria were identical to inclusion criteria of the SDM implementation study [52]. , We included physicians and nurses who worked at one of three departments within the University Cancer Center Hamburg at the University Medical Center Hamburg-Eppendorf during baseline evaluation of the SDM implementation study [52]."

30. Lines 160-163: The recruitment process appears unclear. Is this a convenient sampling where the members of the study team recruit participants while being patients during their physician meetings?
We changed the description of the sample in the revised data collection section to: "Participants were part of a convenience sample of physicians and nurses. Since this is a secondary analysis, inclusion criteria were identical to inclusion criteria of the SDM implementation study [52]. , We included physicians and nurses who worked at one of three departments within the University Cancer Center Hamburg at the University Medical Center Hamburg-Eppendorf during baseline evaluation of the SDM implementation study [52]."

Data analysis

31. Lines 166-167: The authors are invited to specify the underlying rational for establishing the % cut-off of missing items to decide the exclusion (or not) of cases and why they would include them as well as for the analysis of acceptance.
Thank you for your suggestion, we added a reference which recommend a 30% cut off: "Bannon WJ. Missing data within a quantitative research study: How to assess it, treat it, and why you should care. J Am Assoc Nurse Pr. 2015;27:230–2". For reasons for the inclusion of missing cases in completition rate measures, see our answer on your comment 37.

32. Lines 200-205: The paragraph "Patient and public involvement" is written with a different police style. Also, since this information is about the recruitment process of the study, I suggested to the authors to present it previously (e.g., after line 163).
Thank you, we changed the font and moved the paragraph according to your suggestion.

Results

33. Line 218: Has the duration of cognitive interviews been calculated? If so, the authors could present the mean duration and the standard deviation.
Since other measures besides the ORIC were tested during the cognitive interviews, too (see methods section), we decided against details on mean duration / standard deviation of the

interviews, because we do not believe that they would be meaningful for the research question of this study.

34. Line 242: The authors are invited to standardize the unit to which the percentages are presented in the Table 2 (e.g., % of age [<30 years] and % of gender [different gender or preferred not to answer questions]).
Thank you for the suggestion, we changed the units to one decimal place in the revised methods and discussion section (tables and values mentioned in the text).

35. Lines 269-270: The authors report missing data for several items of the questionnaire. Could the authors discuss about the implications related to the data collection procedures used?
Thank you for this suggestion, we added this discussion to the revised discussion section about item: "Since the ORIC was presented as the last measure in a three-page survey, missing values might be a flag for respondent fatigue. However, missing value rates for single items and the overall measure were quite low and the German ORIC was found to be a well-accepted measure." As already discussed in the discussion section, missing values for item 10 might also be due to reduced comprehensibility of this item: "Feedback by participants, members of the study team, and external experts as well as completion rates suggest that item 10 could not be translated satisfyingly and comprehensibility of this item seems to be low [50]."

36. Line 274: Same comment than the one made for line 44: I would suggest to the authors to specifically refer to the internal consistency considering this is the only aspect of reliability that has been studied.
Thank you, we changed it in the whole revised manuscript.

37. Line 279: At the end of the Table 5, the authors specified: "For calculation of response rate, four additional cases were included because these participants only skipped the ORIC but filled out the rest of the questionnaire." This specification could be relevant for future publications with the data collected in the SDM implementation study. However, for this present paper focusing specifically on the ORIC, I do not understand the relevance of including such data.
We think, when participants refused to fill out most of the measure or, like in our cases, the entire measure, it is an important statement for completion rate and acceptance and should be visible in the completion rates. Therefore, we included these cases in calculations of the completition rate. We added this explanation to the revised methods section: "For analysis of acceptance, we also included cases with more than 30% of ORIC items missing because these values are part of completion rate and relevant for interpretation of acceptance." Besides that, the number of "four additional cases" mentioned under Table 5 was a mistake, indeed there were five cases included, as correctly written in the results section for the psychometric evaluation.

Discussion

38. Lines 281-285: Even though the translation process constitutes the first objective of the paper and has been detailed in the Methods section, no paragraph about this process is presented at the beginning of this present section. As stated previously, the authors are therefore invited to detail the results of the translation process and triangulated them (strengths/limitations). *Thank you for the suggestion, we elaborated the revised discussions section and added information about items: "Items 1 to 9 could be translated and adapted successfully after two rounds of cognitive interviews and several rounds of discussions within the study team and with external experts. The translation team found fast consent for items 2 to 5 and 8. These items were also well understood by all participants within the first round of cognitive interviews. For items 1, 6, 7 and 9, translations were less congruent and several adaptations and discussions were necessary."*

39. References

Thank for the detailed reading, we corrected the missing/wrong points below in the revised reference section.

Line 437: The volume of the reference is missing.

Line 448: The volume of the reference is double (123-124) and there is a letter at the end (A).

Line 458: The volume of the reference is double (123-124).

Lines 464-465: The title of this reference is the only one presented in brackets.

Line 479: The place of publication of this reference is missing.

Line 521: The page numbers of this reference are missing.

Lines 527 and 538-539: Some information for these references are missing (publication house and place).

Line 547: The volume and the page numbers of the reference are missing.

Finally, I would like to reiterate my appreciation for this publication. I hope that my comments will be useful for the improvement of this manuscript. The authors will contribute to an important field of research in knowledge translation.

Thank you very much!

Yours sincerely,

MR

---------------------------------RESPONSE TO COMMENTS OF REVIEWER 2 --

Reviewer: 2
Reviewer Name: Mesfin Kassaye Tessma
Institution and Country: Karolinska Institutet, Sweden
Please state any competing interests or state 'None declared': None

Please elaborate no answers
Abstract: It is not accurate, detail comments given.
Objective: Yes in the "Introduction" part.
Study design: It is based on secondary analysis of data. But the details are unclear; requires more elaboration.
References: Some references were not appropriate (please see my comments on "good validity")
Results: To some extent (please see my comments under "General" and "Results"). The results of the CFA cannot be presented as sufficient evidence unless the assumptions are checked and the validity of the findings "confirmed".
Discussion and conclusion: To some extent but the results were not clearly related to construct validity. It requires some elaboration.
Conflicts: I have no idea about "undeclared conflicts"

Language: The language may require editing, to make it appropriate for its intended audience.

As all comments above are described in more detail in the specific comments below, we only responded to the comments below.

Mesfin Tessma Title: "Translation, adaption and psychometric testing of the German version of the Organizational Readiness for Implementing Change measure".

General: The purpose of the study is interesting and relevant and the topic is suitable for BMJ open. This manuscript would therefore come under the remit of this journal. It provides important insight into the topic of ORIC considering the context in Germany. The overall organization of the paper is good and in general the work is a serious and sincere research effort. But it could be improved by the changes recommended in the details. The language may require editing, to make it appropriate for its intended audience.

Thank you for your feedback and your comments!

To Editorial Board

1. The major problem for me in the data analysis is the use of the classical test theory (CTT) when the modern test theory (MTT) is appropriate. To be specific the authors should have used Rasch analysis instead of EFA-CFA to evaluate the psychometric properties. CTT assume interval data and that all items in a scale are equally difficult and does not allow for a separation of persons and items. On the contrary, the Rasch based analysis is used to analyze ordinal data in order to provide linear measures, estimates each item's difficulty as well as each person's ability on the same metric, and examines each item's relationship to the measured theoretical construct. It generates reliability and validity estimates of both persons and items that are independent of the sample distribution.
   Thank you for this comment. CTT and more sophisticated methods like item response theory (IRT) or Rasch analysis (RMT) are different methodologies providing different information. Petrillo et al. (2015) compared the three methods and found similar results for the three methods, with IRT and RMT providing more details on the data structure. CTT is an elaborated psychometric method with several strength like familiarity, ease of adoption and use. It provides tangible statistics that can be checked by comparison to existing criteria. On the other hand, IRT and RMT require complex and advanced levels of mathematical skills and unique software, even if they come along with some advantaged. Petrillo et al. concluded that the researcher should select a method they are comfortable performing and also interests of the intended audience should be considered. Since we do not have a high-stake situation like developing a PRO measure, a diagnostic tool or a measure for education, using CTT for our psychometric evaluation is appropriate.
   Petrillo, J., Cano, S., McLeod, L. et al. (2015). Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. Value in Health, 18:25-34.

2. The results of the CFA cannot be presented as sufficient evidence unless the assumptions are checked and the validity of the findings "confirmed". Therefore, to support the findings it is important, the authors check the statistical assumptions and improve the statistical methods and the result sections.
   For this suggestion, please see our answer on your comment 21.

<u>Title</u>:

3. The title may be modified. "Translation, adaption and psychometric evaluation of the German ..."

Thank you, we modified the title to "Translation and psychometric evaluation of the German version of the Organizational Readiness for Implementing Change measure (ORIC) – a cross-sectional study."

<u>Abstract:</u>

4. Objective: Please check the objectives stated in the text (page 4, line 107). It is not consistent.
Thank you, we checked the consistency. The objective is worded as "to translate the measure Organizational Readiness for Implementing Change (ORIC) into German and assess its psychometric properties." in both the abstract and the introduction.

5. Design: It is unclear and not specific. Please use a better description (secondary analysis of …)
We changed the description of the design to "cross-sectional study" in the revised title, the revised abstract and the revised methods section.

6. Primary and Secondary Outcome Measures: This part included data analysis (not consistent with the subtitle). Please modify the subtitle or the content.
*According to submission guidelines of BMJ Open, this subsection should include: **"primary and secondary outcome measures: planned (i.e. in the protocol) and those finally measured (if different, explain why)".** Accordingly, we described the measures we used for our data analysis. To avoid redundancy, we omit the sentence "For psychometric evaluation we conducted a secondary analysis of baseline data from a SDM implementation study." In the revised abstract.*

7. SDM (?) – please specify.
We specified the abbreviation SDM when first mentioned in the revised abstract by adding "(shared decision-making)".

8. Results: CFA is mentioned as a statistical method in the abstract but in the "Results" part of the abstract results related to CFA were not given. Results from EFA do not provide evidence for construct validity. I suggest to include results from the CFA (if "the German ORIC can be used to analyse organizational readiness for change" – as the authors recommended in the conclusion part).
We specified in the revised abstract that EFA as well as CFA provide evidence for the one-factorial structure of the German ORIC: "EFA and CFA provided a one factorial structure."

9. The finding "low comprehensibility" is unclear. In the text it is defined but difficult to understand in the abstract. This applies also to structural validity.
Thank you for this suggestion, we specified this in the revised abstract: "Translation and cognitive testing of the ORIC was successful except for item 10, which showed low comprehensibility as part of content validity in cognitive interviews." We also specified results of factor analysis: "EFA and CFA provided a one factorial structure."

10. Item difficulty is often used in Rasch analysis not in factor analysis. It can be confusing. It seems as if the authors have performed Rasch analysis.
Calculating item difficulties was not explicitly part of the factor analysis, but was conducted as part of the psychometric evaluation. Item difficulty is not used only in Rasch models, but also is an aspect commonly used in CTT to describe item characteristics or to inform item selection (see for example Boateng et al. "Best practices for developing and validating scales for health, social, and behavioral research: a primer." Frontiers in public health 6 (2018): 149.).

11. Conclusion: The phrase "highly accepted" may be misleading.
Thank you for your comment, we specified this point by changing "highly accepted […] measure" to "measure with high completition rates".

Introduction

12. The use of the terms "good reliability" and "good validity" (page 4, lines 104-105) and the "values" related to the definitions of these terms are often confusing. I don't think the references indicated support the use of these terms and their "acceptable values". Please modify.
Thank you, we modified this part in the revised introduction to: "The ORIC has been psychometrically tested, revealing a completion rate of > 72%, a Cronbach's α of > .80 and two correlating factors."

Methods

13. Subtitle "Psychometric evaluation" (page 9, line 234): "psychometric evaluation" is a better description (in the title for example, it is "psychometric testing")
Thank you, we also think so and changed the title to "Translation and psychometric evaluation of the German version of the Organizational Readiness for Implementing Change measure (ORIC) – a cross-sectional study." We also changed the subtitles of the revised methods and results section to "psychometric evaluation".

Data analysis

14. Missed values (page 6, line 167): The authors have stated that cases were excluded if more than 30% of the ORIC items were missing. Why 30%? They have also replaced the item mean for all missed values (but often multiple imputation is used). Why not multiple imputation instead of mean? The issue of missing values is not as such a problem, if Rasch analysis was used. See my comments under "General".
The exclusion of cases if more than 30% of the values are missing, is recommended by Bannon (2015). We added that reference to the revised methods section. Besides multiple imputation, replacement of means is also a suitable method for our data, so we decided to use this method.

15. Page 6, line 175: "Because the two-factor model could not be confirmed, the data was randomized and split into two subsets (each n =115)". This is very confusing (please see the results of the CFA in Table 4). What does it mean in the light of Table 4? What does it mean that the data set was randomized? How was the randomization process performed? Who performed it? What was the purpose of the randomization?
It is recommended to not calculate EFA and CFA with the same data set. So we used half of the data set for each calculation. AL randomized the participants ID via Excel. We specified information about this process to the revised methods section: "Because the two-factor model could not be confirmed, we decided to calculate an exploratory factor analysis (EFA) and afterwards an additional CFA to check for model fit. It is recommended to not calculate EFA and CFA with the same data so the data set was randomized by AL and split into two subsets. The first 115 randomized cases including all data of participants were added to EFA, the second 115 cases were added to CFA." Because in SPSS the sequence of cases was connected with the date when we received the survey, randomization was necessary.

16. Page 7, line 190: "Because of low content validity of item 10 …" – this is unclear.
Thank you for your comment, we specified that point: "During the translation process and cognitive interviews we found low content validity for item 10 (see results section), therefore the use of item 10 for the German ORIC needs to be evaluated."

17. The authors have used both EFA and CFA, which is the right approach when using CTT. However, EFA helps in evaluating the dimensionality of scales but it does not provide evidence for construct validity. It is CFR that provides evidence for construct validity. As we know construct validity is the extent to which an instrument measures the construct that it is intended to measure.

You are right. But as far as we see, we did not argue that our EFA analysis provides evidence for construct validity. Accordingly, we avoided the term "construct validity" in this context.

18. Please provide information on how the sample size was determined?
    We specified this in the revised data collection section: "Participants were part of a convenience sample of physicians and nurses. Since this is a secondary analysis, inclusion criteria were identical to inclusion criteria of the SDM implementation study [52]. , We included physicians and nurses who worked at one of three departments within the University Cancer Center Hamburg at the University Medical Center Hamburg-Eppendorf during baseline evaluation of the SDM implementation study [52]."

19. The authors did not use parallel analysis to determine the number of factors to be retained in the EFA. Why?
    Thank you for this interesting remark. Indeed, parallel analysis has been shown to be superior to the Kaiser criterion. Therefore, we redid the calculations which again lead to a one-factor solution. Further we revised the method section: "An EFA with oblique rotation was calculated for the first subset. The non-orthogonal rotation was chosen according to Weiner et al. [40]. In their theory, organizational readiness for change consists of two interrelated dimensions, therefore the two factors are expected to be correlated. Analogue to analysis done by authors of the English ORIC [43,61], we extracted components based on parallel analysis. The criterion of parallel analysis was shown to be superior to other statistic criteria like the Kaiser criterion [62,63]. It compares the eigenvalues of the data to the eigenvalues based on random data with equivalent sample size and number of variables and chooses only factors with eigenvalues higher than for random data [61,62]." and the results section: "As shown in table 3, only the first component had an eigenvalue higher than 95% percentile of the eigenvalues of corresponding random data. Thus, according to parallel analysis, a one-factor model was assumed." We added the eigenvalues of random data which are used for parallel analysis to table 3 (and Supplementary File 4 for calculations of the 9-item ORIC version).

Results

20. Table 2: Percentage are given as one or two decimal places. Please round it to one decimal place.
    Thank you for your suggestion, we changed the percentages to one decimal place in the tables and the text of the manuscript.

21. Do the data meet the appropriate assumptions for EFA and CFA?
    There are several assumptions for factor analysis that we checked. The sample size must be large enough. According to the COSMIN criteria the sample size is very good if it is at least "7 times the number of items and ≥100" according to COSMIN criteria. The splitted data set which was used for EFA and CFA had a sample size of n=115. In addition, there were no outliers and we checked with histograms that responses were approximately normally distributed. Also according to our used checklist for reporting standards (see Supplementary File 5), assumptions for EFA were met: "EFA communalities are all above .70, therefore the sample size of n=115 can be determined as adequate." Also inter-item correlation should be sufficient to calculate common factors (each item should have some correlation above 0.3), while at the same time correlations should be low enough (< 0.9) to rule out multicollinearity. As reported in the first draft: "Inter-item correlations ranged from .434 (item2/item9) to .723 (item3/item 5) (see Table C of Supplementary File 1)." Furthermore, we already provided information on Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy Bartlett`s test of sphericity in the first draft: "KMO measure was .933 and Barlett´s test of sphericity yielded $X2 = 1485.11$, $p < .001$, indicating that a factor analysis of the data was appropriate [57,58]". We added more information about assumptions for EFA and CFA to the revised results section: "Sample size was large enough

(>100 [Quelle]) to allow factor analysis, even for a split data set with n=115. Furthermore, there were no outliers and values were approximately normally distributed."

22. Table 3: Please expand the title of the table – the title identifies the data displayed in the table and the context in which they should be interpreted. Many horizontal lines are used in the table. Do we need them?
We expanded the title of table 3 to "Results of EFA with oblique rotation and parallel analysis: eigenvalues of the ten components of the German ORIC and eigenvalues for corresponding random data.". We also deleted the horizontal lines.

23. Why was half of the dataset used for EFA? (Please see footnote)
Please see my answer to your comment 12.

24. How does missing values and outliers affect the result? Please comment!
We did not observe any outliers and only very few missing values. We analysed and reported missing values (see results section) but we don´t think that they affect the results much. Reasons for missing values are discussed in the revised discussion section: "Since the ORIC was presented as the last measure in a three-page survey, missing values might be a flag for respondent fatigue. However, missing value rates for single items and the overall measure were quite low and the German ORIC was found to be a well-accepted measure."

25. Based on AIC and PNFI results the authors preferred the one factor model, although the two factor model has acceptable and better values for TLI and RMSEA. However, since CFA is a confirmatory technique and it is theory driven, it is very important to relate it to the theoretical model. Therefore, the analysis considers the theoretical relationships among the observed and unobserved variables (Schrebier et al, 2006). Thus it seems the two factor model may be better in the light of the theory.
Thank you for your comment. For the original English version of the ORIC, the authors hypothesized two factors based on Weiners theory. Accordingly, when calculating a CFA, they decided for a model with two factors, even they correlate with .56 to .60. In our CFA for our sample we found much higher correlations of 0.87 and EFA provided a clear one-factorial structure. We calculated the model fits to validate the results of the EFA and found good model fits for the more parsimonious one-factor model. When taking all these results into account, we decided for the one-factor model even though it is different to theory. We detailed this discussion this in the revised discussion section: "Because Shea et al. described correlations between the two theory-based factors "change commitment" and "change efficacy" of .56 to .60, we a priory hypothesized a two-factorial structure of the German ORIC. Indeed, we found much higher factor correlations of 0.87. Results of EFA also clearly indicated a one-factor structure, therefore could not confirm the two-factor structure of the English and the translated Danish and French versions of ORIC [43,45,46]. When calculating model fits, the two-factor model showed better fit indices […]."

26. Construct validity is achieved when the measure reflects the framework hypothesized in a hypothesis testing study, adequately establishes that the measurement model fits the actual data. Have the authors considered the theory? The results of the CFA cannot be presented as sufficient evidence unless the assumptions are checked and the validity of the findings "confirmed". Therefore, to support the findings please check and report about the statistical assumptions of CFA. Please provide information about the assumptions of CFA.
Thank you for your comment, you are right about your comments concerning construct validity. Nevertheless, we did not evaluate construct validity and did not test hypotheses about construct validity in this study. For assumptions of CFA, see comment number 21.

<u>Discussion</u>

27. Page 13, line 303: Structural validity "we could not confirm the two-factor structure of the English ..." But the findings indicated the absolute and comparative indices and RMSEA indicated that the two factor model is acceptable. Why did the authors reported that they "could not confirm" the two factor model? EFA explores the factor structure and CFA confirms the factor structure extracted in EFA. Page 13, line 308: "… one factor model should be preferred". It is a very strong statement considering the findings. Please modify (please see also my comments above)

Thank you, we discussed the results of EFA and CFA and reasons for and against the two-factor model (see also our answer on your comment 25). We concluded to prefer the one factor model but changed the wording of the according statement: "Therefore, we prefer the more parsimonious one-factor model."

--------------------------------RESPONSE TO COMMENTS OF REVIEWER 3 ---------------------------------
------

Reviewer: 3
Reviewer Name: Kira Hower
Institution and Country: University of Cologne, Germany
Please state any competing interests or state 'None declared': None declared

Your study provides an important practical contribution to the assessment and examination of organizational readiness for change and is therefore highly relevant for implementation research. Overall, I consider the manuscript to be worth publishing and recommend some adaptations or additions.

Thank you very much for your positive feedback and your suggestions and comments.

Manuscript ID bmjopen-2019-034380
I had the opportunity to review the manuscript entitled "Translation, adaption and psychometric testing of the German version of the Organizational Readiness for Implementing Change measure (ORIC)". This study took advantage of a secondary analysis of data gathered in a SDM implementation study in a German hospital to assess the psychometric properties of the German-language measure "ORIC".
The authors reported that an abbreviated measure reflected the conceptual structure of the construct, convergent validity, as well as item characteristics and reliability. The study has several strengths including that the analytic approach was well-described and appropriate, with an interesting qualitative approach to explore comprehensibility. Unfortunately, the study has some limitations including the cross-sectional nature of the study design, no testing of the criterion-based validity, the fact that the instrument is only applied in one organizational setting and shared method variance as well as the likelihood of halo effect.

Thank you for your summary of strengths and limitation of our study. We elaborated the discussion of the mentioned points in the revised discussion section: "This study has some limitations. First, several psychometric parameters are not analysable because of the secondary analysis and the cross-sectional design of the study. It was not possible to calculate e.g. convergent or divergent validity yet. Second, we applied the ORIC only in three departments of one University Medical Center in Germany. Further validation in different organizational settings is needed to ensure generalizability. Third, for this psychometric evaluation we used a German ORIC, which we adapted and specified for the context of SDM implementation. Our results might not be generalizable for other interventions in other organizations. Fourth, although SDM was not implemented to the participating clinics before, there might be participants who were more familiar with the concept of SDM than others. Fifth, item 10 was again slightly changed after finishing cognitive interviews. This item was not finally tested for comprehensiveness. A major strength of this study is that we provided the first measure to assess organizational readiness for change in German language for use in implementation studies. We conducted an elaborated translation procedure, which was recommended for survey translations. We furthermore used a qualitative approach to explore comprehensibility including discussions with

international colleagues and experts outside of the study team. Furthermore, we assessed the ORIC in a sample including physicians and nurses which was large enough to robustly perform the psychometric analysis on the German version of the ORIC measure."

Introduction
The introduction is well structured. However, a few questions have occurred regarding implementation process, status and outcomes.

1. Has SDM already been implemented in the hospital? From the description of the instrument it can be concluded that there are variations, i.e. SDM may or may not be implemented and known. This should be explained in more detail and further information on the implementation should be provided. If SDM is already implemented, there is a high likelihood of halo effect from this design, i.e., respondents recalling the initiative of SDM implementation and inferring responses to their attitudes to the change based on the outcome. This is a huge problem for retrospective studies that ask respondents to evaluate factors contributing to an outcome that is known.
Thank you for your question. As written in the methods section, only data of the baseline assessment phase of the SDM implementation study were included in the psychometric analyses at hand. That means, our SDM implementation program was not rolled out at the point of assessment. Nevertheless, there might be clinicians, who have heard about the concept of SDM before and others might not. According to your suggestion, we added this concern to the discussions section as part of the limitations: "Fourth, although SDM was not implemented to the participating clinics before, there might be participants who were more familiar with the concept of SDM than others."

2. The wording "implementation of change" can be confusing, since it is not change that is implemented, but a new practice, structure, etc., that requires changes at different levels. Which changes are required by SDM in hospitals could be explained in more detail. This would make it easier to understand how organizational readiness is affected.
Thank you for your comment. The term "implementation of change" is used in the original English ORIC measure and is also theoretically based by Weiner et al. The measure is designed as a generic measure to assess different kinds of "changes" (e.g. new skills), hence the unspecific denomination. SDM is only one example, the one we use within our implementation study. What SDM entails is defined in the revised introduction section of the manuscript: "SDM can be described as an interactional process on the basis of information exchange. Patients and healthcare professionals (HCPs) are equally and actively involved and jointly responsible for the decision [12–14]. This is especially important in situations with complex treatment options and high impact on patients' quality of life [15]. Patients want to be actively involved in decision-making [16] and benefit from SDM by having better knowledge about their disease and treatment, better risk perception and less insecurity and decisional conflict [17,18]." However, the use of the measure is not restricted to SDM.

3. If SDM is already implemented, is there any information on the implementation status that could be related to organizational readiness for change?
Please see our answer to comment 1.

Methods
The methodological and statistical procedures were explained in a clear and understandable way and are in line with the standards.

4. Linking the data with other data, e.g. on the status of implementation or other external criteria, would have significantly increased the validity of the findings and would have contributed a clear added value to the research.
Thank you, we agree that the comparison to external criteria would be interesting. However, this was not possible with the data set at hand and will have to be done in consecutive evaluations.

A few remaining clarifications should be made.

5. Have you tested for differences between groups/personal characteristics/departments? Is there evidence in the literature that there are differences?
    Thank you for your suggestion. We did not test for this within this secondary analysis. Since we only evaluated baseline data before SDM implementation, we had no information about the actual status of SDM adoption in the clinics and therefore could not make hypotheses about group differences. Additionally, we did not use other measures in our survey, which can be used as a gold standard for the assessment of readiness for change and could compared to the new German measure ORIC. So, group differences were not the focus of this publication aiming to psychometrically evaluate acceptability, item analysis, reliability and factorial structure of the ORIC.

6. Do the participants hold leading positions? Has this been taken into account in the survey and analysis? Organizational readiness is clearly influenced by the role in the organization, so this would be, if possible, interesting information to consider.
    Yes, in the psychometric analysis we could include participants of different professions, age and work experiences, including senior and head physicians of the clinics (see sample characteristics in the methods section). So, our data represent different hierarchy levels. We agree that participants' views on organizational readiness might be connected to their status and role in the organization. As already explained in comment 5 we did not generate hypotheses about group differences. This interesting topic was beyond the scope of this paper and analysis.

7. For how many cases were missing values replaced?
    Thank you for this question. You can find this information in the results section: "For items 1 to 9 four to six missing values could be detected. For item 10, nine missing values were found. Taking all items into account, more than 97% of the measure were answered." The missing values can also be extracted from Table 5, which show completition rates for each item.

8. Did the authors make model modifications, such as the correlation of error terms?
    No, we did not conduct model modifications. In the original validation study of the English ORIC, model modifications like correlations of error terms were not calculated and we do not see reasons for calculating model modifications with regards to the content of the items.

Results
The presentation of results is detailed, well structured, and clearly illustrated. Regarding the results themselves I have the following minor concerns that may need to be addressed.

9. How do the authors explain the fact that in this study, in contrast to the other validation studies, a single factor structure of the scale emerges? What does this mean in terms of a limitation? *Thank you. This is an interesting question. Despite CFA and EFA provided a clear one-dimensional structure, a one-factorial model did not clearly fit on the data. We added this topic to the revised discussions section: "These differences compared to previous validation studies might be a consequence of diverse cultural connotations of the ORIC items in different languages, the adaption to the context of SDM or might be due to specific characteristics of the participating clinics." But of course, further evaluation of the ORIC with focus on other kinds of changes would be helpful to get a better picture of these relations.*

10. Also, knowledge about SDM and the organization must be available to answer the questions. – *Yes, you are right. As you can see in the section about cognitive interviews in the methods part, we provided a definition of SDM in the introduction of the survey. We also developed an introductory description of the ORIC, which motivated participants to think about the clinic, they are working in, when answering the item (see Supplementary File)."*

11. Since the descriptive statistics show that the average values are rather in the middle, it might be that the participants did not know enough to answer the questions? How could this be ensured and would this be an important prerequisite in terms of using the scale?
Thank you for your question. Since we provided a definition of SDM in the introduction of the survey, participants should at least have an idea about SDM. But because they filled out the ORIC before SDM implementation in their clinic, they might not be aware of all consequences of SDM implementation. Therefore, we would expect change sensitivity of the ORIC (which we could not test within this study). Nevertheless, we think it is reasonable to ask healthcare professionals about their view on readiness for implementing SDM in their clinic already before implementation because this might be an important barrier or facilitator for implementation success.

Discussion
The discussion is well structured, but is brief and could benefit from more explanations and implications.

12. The limitations could be extended, e.g. with regard to the points mentioned above.
Thank you for the suggestion, we expanded the strength and limitations part and added some limitations: "Third, for this psychometric evaluation we used a German ORIC, which we adapted and specified for the context of SDM implementation. Our results might not be generalizable for other interventions in other organizations. Fourth, although SDM was not implemented to the participating clinics before, there might be participants who were more familiar with the concept of SDM than others. Fifth, item 10 was again slightly changed after finishing cognitive interviews. This item was not finally tested for comprehensiveness."

13. In relation to the conclusion, the authors could make it clearer what the theoretical and practical contribution of the study is and what the results contribute to research.
Thank you for the suggestion, in the revised Conclusion section we added some suggestions for future studies: "To increase content validity of the measure, the use of a 9-item German ORIC (without item 10) should be evaluated in future studies."

VERSION 2 – REVIEW

REVIEWER	Mélanie Ruest Université de Sherbrooke, Canada
REVIEW RETURNED	19-Mar-2020

| GENERAL COMMENTS | Thank you for the opportunity to review this paper a second time. I carefully consider the explanations provided and take note of the changes made to the manuscript. I have chosen to provide free-text comments for each section of the paper, with the reference to specific lines when it was relevant. However, I add some comments to this end in the suggestions below.

Overall advice
Given the authors responded with satisfaction to the questions, I recommend the acceptance of this manuscript. Nevertheless, I use this second round of revision to present below minor (mainly linguistic) suggestions for additional improvements. I am not qualified to assess the quality of the language in this paper since English is my second language. However, I made a throughout reading of the manuscript and I invite the authors to revise a second time the language of the manuscript since several errors have been targeted. |

Free-text comments/suggestions
Abstract

Objectives
Line 32: The order of the following words "measure" and "Organizational Readiness for Implementing Change" differs from one place to another throughout the manuscript (lines 66, 109, 165, 183, 258, 320, 388). Please standardize as presented at line 66.

Design
Line 35: If the abbreviation SDM is not used in other sections of the abstract, it could be removed since it is well-presented at line 81 in the introduction.

Line 38: The number of participants is usually specified in brackets (e.g., healthcare professionals (HCPs; n=11 and HCPs; n=230).

Results
Line 45: If possible, specify the translation of the ORIC to which the authors refer (e.g., Translation and cognitive testing of the German ORIC […]).

Line 53: The terms "To analyze" and "To analyse" are interchangeably used. The authors are invited to uniformize it throughout the manuscript.
For the verb "To analyze", see lines 53, 101
For the verb "To analyse", see lines 290 and 399 and Supplementary File 4 (line 19)

Free-text comments/suggestions
Manuscript

Introduction
Lines 87-88: Please review the structure of the examples given to clarify the levels of the organization.
Here's a suggestion: (i.e., individual, group, organizational and system levels).

Line 106: The following characters "of >" could be replaced by "superior to".

Methods
Translation
The translation process is well-detailed and allow to better illustrate the steps accomplished for developing the German version of the ORIC. Here some additional comments/questions to bonify this paragraph.

Line 125: The authors are invited to add reference(s) about their statement on the growing recognition of the TRAPD protocol in the field of translation.

Lines 136-137: I read about the TRAPD protocol of translation to better understand the different steps presented in this paragraph and I have been able to situate each of them with the explanations given, excepted for the Pretesting phase.
Did the authors perform this step? They are invited to clarify this aspect.

Line 139: The terms "adaptation" and "adaption" are interchangeably used. The authors are invited to uniformize it throughout the manuscript.
For the term "adaptation", see lines 184, 241, 335, 470
For the term "adaption", see line 353 and Supplementary File 5

Patient and public involvement
Lines 183-185: The title of the paragraph "Patient and public involvement" should be revised (unless it's a rule of the journal) as the authors specify at the end of this brief section that no patients were involved in this study. Also, the structure of the first sentence should be revised.
Here's a suggestion: "The ORIC measure preliminary addresses HCPs. Physicians, nurses and psychooncologists were therefore involved in the adaptation of the measure […]".

Finally, in its current form, this paragraph contains information already specified in preceding sections of the manuscript. I tend to no longer really see the relevance of this paragraph.

Data analyses
Lines 219-220: The authors are invited to review the structure of this sentence.
Here's a suggestion: "During the translation process and cognitive interviews, we found low content validity for item 10 (see results section). therefore The use of item 10 for the German ORIC needs therefore to be evaluated."

Results
Translation
Lines 233-234: There is a mistake at the end of the sentence.
Here's a suggestion: "[…] in their translations of items 2 to 5 and 8 ".

Line 234: The authors are invited to specify what they intend by "Greater translation differences". Even though the authors added relevant details in the Results section about the translation process, additional information could be presented to this end. How many differences have been found for the translation step? Were they similar or different among the items 1, 6, 7, 9 and 10? Could the authors give examples of differences found in sentence structure or single words?

Line 246 (same comment than the one made for lines 233-234): There is a grammatical error for the word item.
Here's a suggestion: "[…] as well as for items 2 to 5 and 8".

Line 255: There is a missing letter at the beginning of the following sentence and it could be better phrased.
Here's a suggestion: "So In a next step, we consulted with DF and reached […]".

Factor analysis
Line 277: Please review the structure of the following sentence: "Furthermore, there were no outliers and values were approximately normally distributed."

Line 300: There is a mistake at the end of the sentence: "[…] (see Supplementary File 4, Tables A and B)."

| | Line 307: A comma is missing at the beginning of the sentence: "For items 1 to 9, between four and six missing values […]."

Line 310: Please review the punctuation sign used for the items: "[…] and inter-item corrlations from .434 (items 2 to 9) […]" OR "[…] and inter-item correlations from .434 (items 2-9) […]"

Discussion
Line 325: Please review the structure of the following sentence: "Items 1 to 9 have been translated and adapted successfully after two rounds of cognitive interviews […]."

Strength and Limitations
Line 372: I see minimally two strengths in this section (i.e., elaborated translation procedure and psychometric assessment of the German version of the ORIC). I would therefore tend to present this section as "Strengths and Limitations" |

REVIEWER	Mesfin Tessma Karolinska Institutet, Sweden
REVIEW RETURNED	15-Mar-2020

GENERAL COMMENTS	Most of my comments and suggestions are incorporated.

REVIEWER	Dr. Kira Isabel Hower University of Cologne, Germany
REVIEW RETURNED	03-Mar-2020

GENERAL COMMENTS	The authors have sufficiently addressed the critical issues and suggestions I have made. The manuscript has benefited greatly from the overall editing and has improved considerably. Despite the limitations of the study due to the design and the available data, I recommend the manuscript for publication. This approach of validating the scale allows a valuable groundwork for addressing the remaining aspects of validation.

VERSION 2 – AUTHOR RESPONSE

Response to reviewer 1

Reviewer Name: Mélanie Ruest
Institution and Country: Université de Sherbrooke, Canada
Please state any competing interests or state 'None declared':
As already stated in the previous revision:
No competing interests to declare. Please note that one of my published studies is used in the references of this study. As stated in the manuscript, I exchanged emails with an author to discuss about our respective experiences of translation for an item of the questionnaire.

Please leave your comments for the authors below
March 19, 2020

Thank you for the opportunity to review this paper a second time. I carefully consider the explanations provided and take note of the changes made to the manuscript. I have chosen to provide free-text comments for each section of the paper, with the reference to specific lines when it was relevant. However, I add some comments to this end in the suggestions below.

<u>Overall advice</u>
Given the authors responded with satisfaction to the questions, I recommend the acceptance of this manuscript. Nevertheless, I use this second round of revision to present below minor (mainly linguistic) suggestions for additional improvements. I am not qualified to assess the quality of the language in this paper since English is my second language. However, I made a throughout reading of the manuscript and I invite the authors to revise a second time the language of the manuscript since several errors have been targeted.

➔ *Thank you very much for your positive feedback, your very detailed reading and your suggestions to further improve the manuscript.*

<u>Free-text comments/suggestions: Abstract</u>

Objectives
Line 32: The order of the following words "measure" and "Organizational Readiness for Implementing Change" differs from one place to another throughout the manuscript (lines 66, 109, 165, 183, 258, 320, 388). Please standardize as presented at line 66.

➔ *Thank you for the detailed reading. We changed the word order according to your suggestion in the revised manuscript.*

Design
Line 35: If the abbreviation SDM is not used in other sections of the abstract, it could be removed since it is well-presented at line 81 in the introduction.

➔ *We deleted the abbreviation in the abstract of the revised manuscript.*

Line 38: The number of participants is usually specified in brackets (e.g., healthcare professionals (HCPs; n=11 and HCPs; n=230).

➔ *Thank you for this suggestion. We put the number of participants in brackets: "[…] we conducted cognitive interviews with healthcare professionals (HCPs, n=11). Afterwards, HCPs (n=230) filled out the measure."*

Results
Line 45: If possible, specify the translation of the ORIC to which the authors refer (e.g., Translation and cognitive testing of the German ORIC […]).

➔ *According to your suggestions we changed the sentence to: "Translation and cognitive testing of the German ORIC was successful except for item 10 […]" in the revised abstract.*

Line 53: The terms "To analyze" and "To analyse" are interchangeably used. The authors are invited to uniformize it throughout the manuscript.
For the verb "To analyze", see lines 53, 101
For the verb "To analyse", see lines 290 and 399 and Supplementary File 4 (line 19)

➔ *Thank you very much. We changed the term to the American English term "analyze" in the whole revised manuscript and the Supplementary Files.*

Introduction
Lines 87-88: Please review the structure of the examples given to clarify the levels of the organization.
Here's a suggestion: (i.e., individual, group, organizational and system levels).

➔ *According to your suggestion, we changed the sentence in the revised manuscript to: "When implementing SDM or other interventions in organizations, several barriers on different levels of the organization (i.e., individual level, group level, organizational and system level) need to be considered [1,5–7,10,29–32]."*

Line 106: The following characters "of >" could be replaced by "superior to".

➔ *Thank you, we changed the sentence in the revised manuscript to: "The ORIC has been psychometrically tested, revealing a completion rate of more than 72%, a Cronbach's α of above .80 and two correlating factors [43,45,46]."*

Methods
Translation
The translation process is well-detailed and allow to better illustrate the steps accomplished for developing the German version of the ORIC. Here some additional comments/questions to bonify this paragraph.

➔ *Thank you very much for your positive feedback!*

Line 125: The authors are invited to add reference(s) about their statement on the growing recognition of the TRAPD protocol in the field of translation.

➔ *There are several articles, which also used team translation protocols or compared different translation procedures. We added some references, which provide a broader picture of the field of translation processes:*

49. Forcino RC, Bustamante N, Thompson R, Percac-Lima S, Elwyn G, Pe´rez-Arechaederra D, et al. Developing and Pilot Testing a Spanish Translation of CollaboRATE for Use in the United States. PLoS One. 2016;11:e0168538.

50. Epstein J, Osborne RH, Elsworth GR, Beaton DE, Guillemin F. Cross-cultural adaptation of the Health Education Impact Questionnaire: experimental study showed expert committee, not back-translation, added value. J Clin Epidemiol. 2015;68:360–9.

51. Perneger TV, Leplège A, Etter JF. Cross-cultural adaptation of a psychometric instrument: two methods compared. J Clin Epidemiol. 1999;52:1037–46.

52. Ponce NA, Lavarreda SA, Yen W, Brown ER, DiSogra C, Satter DE. The California Health Interview Survey 2001: translation of a major survey for California's multiethnic population. Public Heal Rep. 2004;119:388–95.

Lines 136-137: I read about the TRAPD protocol of translation to better understand the different steps presented in this paragraph and I have been able to situate each of them with the explanations given, excepted for the Pretesting phase. Did the authors perform this step? They are invited to clarify this aspect.

➔ *According to the work of Harkness and colleagues, pretesting includes techniques for evaluation of survey items before data collection begins. Aim is to test for viability and refine the survey items. We performed pretesting by conducting cognitive interviews with health-*

care professionals and subsequently adapting the scale. We added this information to the revised methods section: "As a next step we pretested the translated measure by conducting cognitive interviews and thereby assessed comprehensibility as part of content validity."

Line 139: The terms "adaptation" and "adaption" are interchangeably used. The authors are invited to uniformize it throughout the manuscript.
For the term "adaptation", see lines 184, 241, 335, 470
For the term "adaption", see line 353 and Supplementary File 5

➔ *Thank you for the detailed reading. We changed the term to "adaptation" in the whole revised manuscript and the Supplementary Files.*

Patient and public involvement
Lines 183-185: The title of the paragraph "Patient and public involvement" should be revised (unless it's a rule of the journal) as the authors specify at the end of this brief section that no patients were involved in this study. Also, the structure of the first sentence should be revised.
Here's a suggestion: "The ORIC measure preliminary addresses HCPs. Physicians, nurses and psychooncologists were therefore involved in the adaptation of the measure […]". Finally, in its current form, this paragraph contains information already specified in preceding sections of the manuscript. I tend to no longer really see the relevance of this paragraph.

➔ *The journal requires a paragraph including statements about patient and public involvement: "To support co production we request authors provide a short Patient and Public Involvement paragraph as a subsection within the methods section of their papers. If patients and the public were not involved in any way this must be formally documented in the Patient and Public Involvement subsection." We therefore leave this statement in the methods section. Additionally, we changed the structure of the first sentence in the revised manuscript according to your suggestion: "Physicians, nurses and psychooncologists were involved in the adaptation of the measure by taking part in cognitive interviews."*

Data analyses
Lines 219-220: The authors are invited to review the structure of this sentence.
Here's a suggestion: "During the translation process and cognitive interviews, we found low content validity for item 10 (see results section). therefore The use of item 10 for the German ORIC needs therefore to be evaluated."

➔ *Thank you for the suggestion, we changed the sentence in the revised manuscript according to your suggestion: "During the translation process and cognitive interviews, we found low content validity for item 10 (see results section). Thus, the use of item 10 for the German ORIC needs to be evaluated."*

Results
Translation
Lines 233-234: There is a mistake at the end of the sentence.
Here's a suggestion: "[…] in their translations of items 2 to 5 and 8 ".

➔ *Thank you. This was a mistake. We changed the sentence to "[…] in their translations of items 2 to 5 and 8 as well as the response scale.".*

Line 234: The authors are invited to specify what they intend by "Greater translation differences". Even though the authors added relevant details in the Results section about the translation process, additional information could be presented to this end.
How many differences have been found for the translation step?

Were they similar or different among the items 1, 6, 7, 9 and 10?
Could the authors give examples of differences found in sentence structure or single words?

→ *Thank you very much for this suggestion. We added the following information to the revised manuscript: "Both translators (AL and SZ) and the reviewer (IS) did not differ much in their translations of items 2 to 5 and 8 as well as the response scale. For these items and the response scale, only the choice of single words differed without differences in meaning. Greater translation differences were found for items 1, 6, 7, 9, and 10. For item 1, the word "committed" was differently translated. For items 6 and 7, differences were found in the translation of the phrase "feel confident" and the sentence structure. For item 9, differences mainly addressed translation of the phrases "feel confident" and "adjust to this change". For item 10, differences occurred in the translation of the term "manage the politics" and the sentence structure."*

Line 246 (same comment than the one made for lines 233-234): There is a grammatical error for the word item. Here's a suggestion: "[…] as well as for items 2 to 5 and 8".

→ *Thank you, in the revised manuscript we changed the sentence according to your suggestion: "[…] as well as for items 2 to 5 and 8".*

Line 255: There is a missing letter at the beginning of the following sentence and it could be better phrased. Here's a suggestion: "So In a next step, we consulted with DF and reached […]".

→ *Thank you, we added the missing letter: ""Thus, in a next step we consulted with DF and reached […]".*

Factor analysis
Line 277: Please review the structure of the following sentence: "Furthermore, there were no outliers and values were approximately normally distributed."

→ *Thank you. In the revised manuscript we changed the sentence according to your suggestion and also changed the structure of the sentence before: "Requirements for a factor analysis were met [50]. Sample size was large enough (>100), even for a split data set with n=115. Furthermore, no outliers were found and data values were approximately normally distributed."*

Line 300: There is a mistake at the end of the sentence: "[…] (see Supplementary File 4, Tables A and B)."

→ *Thank you. We changed the words in brackets to: "[…] (see Supplementary File 4, Tables A and B)."*

Line 307: A comma is missing at the beginning of the sentence: "For items 1 to 9, between four and six missing values […]."

→ *Thank you. We added the comma to the revised manuscript: "For items 1 to 9, between four and six missing values […]."*

Line 310: Please review the punctuation sign used for the items: "[…] and inter-item correlations from .434 (items 2 to 9) […]" OR "[…] and inter-item correlations from .434 (items 2-9) […]"

➔ *Thank you for the suggestions. Since we describe a correlation between two items, we changed the wording to: "and inter-item correlations from .434 (item 2 and item 9) to .723 (item 3 and item 5)" in the revised manuscript.*

Discussion
Line 325: Please review the structure of the following sentence: "Items 1 to 9 have been translated and adapted successfully after two rounds of cognitive interviews […]."

➔ *Thank you for the suggestion. We changed the sentence to: "Items 1 to 9 were translated and adapted successfully after two rounds of cognitive interviews […]." in the revised discussion section.*

Strength and Limitations
Line 372: I see minimally two strengths in this section (i.e., elaborated translation procedure and psychometric assessment of the German version of the ORIC). I would therefore tend to present this section as "Strengths and Limitations"

➔ *We agree. The heading of this paragraph was changed to "Strengths and Limitations" in the revised manuscript.*

Response to reviewer 2

Reviewer Name: Mesfin Tessma
Institution and Country: Karolinska Institut, Sweden
Please state any competing interests or state 'None declared': None

Please leave your comments for the authors below
Most of my comments and suggestions are incorporated.

➔ *Thank you very much for your positive feedback and the recommendation for publication.*

Response to reviewer 3

Reviewer Name: Dr. Kira Isabel Hower
Institution and Country: University of Cologne, Germany
Please state any competing interests or state 'None declared': None declared.

Please leave your comments for the authors below
The authors have sufficiently addressed the critical issues and suggestions I have made. The manuscript has benefited greatly from the overall editing and has improved considerably. Despite the limitations of the study due to the design and the available data, I recommend the manuscript for publication. This approach of validating the scale allows a valuable groundwork for addressing the remaining aspects of validation.

➔ *Thank you very much for the positive feedback and the recommendation for publication.*