Author's response to reviews

**Title:** Adaptation of Chinese and English versions of the Ankylosing Spondylitis Quality of Life (ASQoL) scale for use in Singapore

**Authors:**

Ying Ying Leung (katyccc@hotmail.com; katy.leung.y.y@singhealth.com.sg)

Weixian Lee (lee.weixian@yahoo.com.sg)

Nai Lee Lui (Lui.nai.lee@singhealth.com.sg)

Matthew Rouse (mrouse@galen-research.com)

Stephen McKenna (SMcKenna@galen-research.com)

Julian Thumboo (julian.thumboo@singhealth.com.sg)

**Version:** 1  **Date:** 30 May 2017

**Author’s response to reviews:**

31 May 2017

Editor in-Chief
BMC Musculoskeletal Disorders

Sir,

Re: BMSD-D-17-00196

Adaptation of Chinese and English versions of the Ankylosing Spondylitis Quality of Life (ASQoL) scale for use in Singapore.

Thank you for your letter dated 5 May 2017. Following the Reviewers’ comments and advice, we have revised our original manuscript. We are pleased to re-submit our revised manuscript and our point-by-point responses to the Reviewers’ advice and comments. Changes are highlighted in the revised manuscript.

The revised manuscript has also been approved by all the authors and that they have given necessary attention to ensure the integrity of the work.

Thank you for your kind attention and we look forward to hearing your favorable reply.
Yours sincerely,

Dr. Ying-Ying Leung
Corresponding author
Department of Rheumatology & Immunology
Singapore General Hospital
Tel No.: (65) 63265276
Fax No.: (65) 62203321
E-Mail: katyccc@hotmail.com

Encl.

Point-to-point response to Reviewer(s)' advice and comments

Reviewer 1: Comments to the Author

This study shows data on the adaptation of Chinese and English versions of the Ankylosing Spondylitis Quality of Life (ASQoL) scale for use in Singapore. I agree with the authors that this study addresses an important subject since quality of life, assessed by the Ankylosing Spondylitis Quality of Life (ASQoL) questionnaire, has been increasingly recognised as an essential outcome in axial Spondyloarthritis. It is of special value since it is a relevant patient reported outcome measure (PROM).

The conclusion of the authors that the new version can be used in Singapore for clinical trials and daily practice in patient speaking either Chinese or English, is much to positive. In this study the data of the Chinese version was to weak and further studies will be required to confirm the properties of these Chinese language version. Moreover the sample size for the English version should be sufficient to allow additional assessment of scaling properties.

Reply: We thank the reviewer for comment. We have revised our statistical analysis to include a RASCH model analysis. We hope these response and revisions will be considered a satisfactory improvement.

My major concerns that should be solved are the following

1. Although the initial conduction of this study was well designed, the operational part was not sufficient enough to conclude that both versions can be used. Based on the data of this study I would suggest to present the data for both the English and the Chinese version separately

Reply: We thank the reviewers for comment. Singapore is a multiethnic Asian country, whereby 80% of Singaporean are literate in English and 71% are literate in two or more languages, PROMs available in both English and Chinese will cover 98% of the population. As the PROMs will be used in the Singapore local setting where patients are filling in with either language of
their choice, both clinical trial settings as well as in observational studies, we think evaluating the psychometric properties of both adapted version is a practical and reasonable approach. We have put in additional work in the assessment of scaling properties, including a RASCH model analysis and DIF of language of administration, which should support the robustness of these versions.

Action: Introduction, page 6

“With a population below 7 million it is difficult to justify adapting a measure into the several different languages spoken in the country. Singapore is a multiethnic country, with 80% of Singaporeans literate in English and 71% literate in two or more languages, PROMs available in both English and Chinese will cover 98% of the population (17). Consequently, a decision was taken to produce Chinese and English language versions of the ASQoL and to validate them with a combined sample of Chinese and English speaking AxSpA patients. This is practical and aligned to the daily clinical practice, observational cohorts and clinical trial in the local settings.”

2. The optimal cross-cultural adaptation of the ASQoL involves also scaling properties by using Rash analysis. I would suggest to perform these analysis for the English version since the sample size should be enough. For Rasch analysis, data from 50 respondents allows 99% confidence that the item calibrations are stable within ± 1 logit.

Reply: We have performed a RASCH model analysis on the combined versions. Additional description is added on Methods (page 14-15), results are presented on page 18-19.

Furthermore I have some minor concerns that should be taken into count in a new version Method section;

1. Translation session

a. Describe and give the reference on what theoretical ground the translation and cross-cultural adaptation process was carried out; Was it based on best-practice methodology for ASQoL, including the reference of the dual panel methodology

Reply: this dual translation was based on the best-practice methodology for ASQoL.

“The ASQoL was translated into Singapore Chinese and Singapore English using the dual-panel methodology (18). Studies have shown that the dual-panel methodology produces translations that are more acceptable to patients than the standard forward-backward methodology (19, 20). The methodology involves conducting two independent translation panels; a bilingual panel followed by a lay panel. First, a bilingual panel produces an initial translation of the questionnaire into the target language. The bilingual panel was conducted with presence of the original developer of ASQoL (SM), who advised on the closest meaning of the items and instructions to the original English language wording.”
2. Psychometric validation

a. What kind of patients were excluded?

b. Recruitment was done by consecutive patients in outpatient clinic of tertiary hospital.

Reply: Consecutive patients attended designated Spondyloarthritis clinic were recruited.

Action: Method, page 9

“For the purpose of culturally adapting the ASQoL, we conducted a cross-sectional study with consecutive patients with AxSpA to determine the validity of ASQoL via classical theory testing (internal consistency, convergent validity and known-group validity) and Item Response Theory. We excluded non-residents (non-Singapore permanent residents or citizens), subjects who could not read English or Chinese, and patients who could not give informed consent.”

What is the generalizability of this group of patients compared to the rest of Singapore?

c. Why did they choose to include only patient that fulfilled the ASAS criteria for axSpA? These ASAS criteria do work for trials, but not for daily practice. Therefore the overall aim as mentioned in the introduction should not be for daily practise

Reply: the original ASQoL was derived for ankylosing spondylitis (AS). However, fulfilling the established classification criteria for AS required the presence of sacroiliitis on radiograph, which is typically a late stage of disease. The new ASAS criteria for axial SpA includes patients with inflammatory back pain, together with a few clinical features, or inflammatory signals on magnetic resonance imaging; and therefore more practical for daily clinical practice.

Action: Methods, page 8

“The ASAS criteria for AxSpA allow classification of patients with inflammatory back pain, other SpA features or signs of inflammation on imaging (either Magnetic Resonance Imaging or radiography); which is representative of the cases of AxSpA in daily clinical practice.”

d. The ASQOL was developed for AS, however non-Singapore studies showed that it can be used in axSpA. Are the populations AS and axSpA in Singapore comparable to those US and EUROPEAN countries? If not, I suggest to use the ASQOL only for the AS population

Reply: Population of axSpA in Singapore have been reported to be similar in terms of clinical features compared to Caucasian studies. Thus we think it is reasonable to extend the application to axial SpA.

Action: Methods, page 9

“Clinical features of patients with AxSpA in Singapore have been reported to be similar to those in Caucasian countries (23).”
3. Test-retest reliability

a. Reproducibility is best evaluated by focusing on patients with relatively stable disease conditions over time. Therefore, patients should be excluded from these analyses if they reported significant changes. Was this performed? And if not you should mention this.

b. Describe that values greater than 0.85 for rank correlation coefficients were considered evidence of high reliability.

Reply: We have specified this in the methods session.

Action: Methods, page 15-16

“The Spearman’s rank correlation of >0.85 for rank correlation coefficients were considered low random measurement error and evidence of high reproducibility. (38).”

4. Statistical paragraph

a. Describe for the convergent validity what associations has been made. ASQOL to….. Describe also that moderate correlations were expected, indicating that the scales assess different but related constructs.

Reply: We have stated the hypothesized magnitude of correction expected for ASQoL with other PROMs.

Action: Methods, page 13

“Strong (r ≥ 0.7) and moderate correlation coefficients (r = 0.5-0.7) suggest that the scores from two PROMs are measuring related construct whereas weak correlation coefficients (r ≤ 0.3) suggest the PROMs are measuring different construct (34). We hypothesize the ASQoL should correlate with SF-36 subscales at least moderately, and to a lesser extent with pain or global assessment indexes.”

b. Sample size calculation is missing. Traditional statistical analyses generally employ sample sizes of 40 or more in this kind of studies.

Reply: We added a paragraph to describe the sample size consideration.

Action: Methods, page 12

“Sample size considerations

A larger sample size is usually required for Rasch model analysis compared to classical theory test. Sample size for adequate Rasch model analysis depends on factors such as the discriminative ability of the items, and the number of item parameters that are being estimated. Recommendations suggest seven times the number of items of the PROM of interest, and >100
patients. A sample size of more than 200 patients per subgroup for proper DIF evaluation was suggested (32), which would be an unrealistic target for a low prevalence disease in a single center study. With these considerations, we decided to include at least 100 patients for the cross sectional study. For the test-retest reliability study, we assumed that the expected correlation coefficient would be 0.85 (type I error rate=0.05, type II error rate = 0.200); the minimum sample size was eight patients. We considered 30 patients in total for both languages, taking into account non-response at time 2.”

Results

1. CDI paragraph. Give some more insight in this group about the % AS within the axSpA and what was the but disease duration?

Reply: We presented detailed description of the cohort on Result, page 17. Up to 83% of the cohort fulfilled classification criteria of AS. The disease duration was not collected unfortunately.

Action: Results, page 17

“One hundred and eighty-seven consecutive patients with AxSpA were recruited in the cross-sectional study, of whom 183 gave complete data and 82% completed the English version. Of these, 175 patients had radiographic images available. One-hundred and forty-five (83%) fulfilled the New York modified criteria for sacroiliitis (39). Among the 30 patients with radiographic features not fulfilling these criteria, 20 had active sacroiliitis on magnetic resonance imaging (22).”

2. Psychometric validation describe number of Chinese n= 33 and n=150 English speakers

3. Table 2 ; HIGH SD for the ASQAL. Please provide the range and the % of the maximum and minimum score to see whether there was a possibility ceiling effect for the ASQOL

Reply: We have added columns in table 2 to show the ceiling (19%) and floor effect (0%) of ASQoL. Similarly was described in text. (Results, page 17-18)

4. Cronbach’s alpha coefficient for the ASQoL was 0.88 indicating good internal consistency. How was this for the English version separately?

Reply: The Cronbach’s alpha coefficient for Singapore ASQoL Chinese and English version separately were 0.93 and 0.86. (in Results, page 17-18).

5. Table 3 Describe that these are the Spearman’s rank correlation coefficients

Reply: We have specified these are the Spearman’s rank correlation coefficients both in text (Methods, page 18) and table 3.
6. Test-retest reliability; Only 42 AS patients completed the ASQoL (relative more from the Chinese baseline sample) Compared to the English speaking patient (23 of the baseline) in this sample. Were the non-responders different from the baseline characteristic’s

Reply: The cross sectional study and the test-retest study were separate studies, recruited patients from the same clinic at different period of time. As the primary question of the test-retest study was to address whether ASQoL was reliable while patient’s condition was stable, patients recruited to this study is not consecutive, and not mean to represent the whole cohort of axial SpA. Patients having unstable condition and changing of medication/ exercise intervention were not recruited.

Action: Methods, page 15:

“We only included patients with stable SpA, and excluded those we were expecting a change in condition, such as those having changes in medication regimen or changing exercise intervention or plan.”

Reviewer 2: This is a potentially relevant and interesting study on the translation and validation of the ASQoL for use in Singapore. In its current form, however, I feel that more information needs to be provided on several steps of the study and analyses to comprehensively judge the findings. Also, the overall claim and conclusion drawn currently seem somewhat overly optimistic to me given the several limitations of the current study and the actual results of the (very basic) analyses done.

Abstract

1) The defined objective in the abstract is quite different from the aim described at the end of the introduction, and in my opinion formulated somewhat too ambitious and vague. To establish true equivalence of a translated scale one should use rigorous advanced statistical modeling techniques in appropriate and large samples (e.g., by using multi-group factor analyses and/or IRT) which were not done in this study. Moreover, what does it mean to have "the same psychometric properties", do the authors mean that it should meet (similar) basic or minimum psychometric standards?

Reply: We thank the reviewer for comment. We have revised our statistical analysis to include a RASCH model analysis. We hope these response and revisions will be considered a satisfactory improvement.

Methods

2) Please describe the test conditions of the measurements in the cross-sectional and test-retest study (e.g., paper and pencil questionnaires? Completed independently at home? Instructions? etc.).
Reply: All questionnaires were completed in paper pencil format. Patients in the cross sectional study completed questionnaires in the clinic (Methods, page 10). Patients in the test-retest study completed both sets of questionnaires at home.

Action: Methods, page 15

“We standardized the administrations so that the forms were completed in their rest time at home on each occasion. Two sets of questionnaires in paper and pencil format were given to patients and completed forms sent back via return envelopes.”

Measures

3) Please provide all relevant psychometric properties of the ASQoL found in the original validations. This is essential to interpret and compare the current findings on convergent validity, internal consistency and test-retest reliably.

Reply: We have added the relevant psychometric properties of ASQoL from the original validation.

Action: Methods, page 10

“The original versions had excellent internal consistency (Cronbach’s alpha = 0.89-0.91), test-retest reliability (r=0.92 UK and 0.91 NL), and fulfilled the requirement of unidimensionality of the Rasch model (3).”

4) For all other measures, the authors should briefly describe their psychometric properties, preferably and if relevant, supplemented by internal consistency in the current study.

Reply: We have provided for the main comparator instrument (SF-36) and HAQ, the relevant psychometric properties.

Action: Methods, page 10 and 11

“The Chinese version of HAQ have been proven valid and reliable (test re-test reliability, r=0.84) in rheumatology settings in Singapore (25)”

and

“Both English and Chinese versions of the SF-36 have been validated for use in the general population of Singapore (27). The validity and reliability of SF-36 have been evaluated, with Cronbach’s alpha of subscales ranged from 0.88-0.90 (28).”

5) Did the patients complete the other measures in the same primary language as the ASQoL and were these validated for use in Singapore English and Chinese populations? If so, please provide references or state this more clearly if this was not the case (and add a reflection on this in the discussion).
Reply: Yes, patient completed the other PROMs in the same primary languages as the ASQoL. SF-36, HAQ have been validated in both languages for use in Singapore (as described in Methods session). The other PROMs were not properly validated in the target population. We have addressed this in discussion.

Action: Discussion, page 23

“Of note is that the Bath indices were not properly culturally adapted and validated in the Singaporean languages. However, the main comparator PROMs, SF-36 and HAQ have been properly validated in our population.”

Translation

6) Why exactly is this "dual-panel" methodology better than other traditional forward-backward methodologies (which are still generally considered the gold standard) and how was this established? A reference to one expert opinion paper in the discussion does not seem to suffice here.

Reply: this dual translation was based on the best-practice methodology for ASQoL. The justifications of using the dual panel translation is given below.


“The ASQoL was translated into Singapore Chinese and Singapore English using the dual-panel methodology (18). Studies have shown that the dual-panel methodology produces translations that are more acceptable to patients than the standard forward-backward methodology (19, 20). The methodology involves conducting two independent translation panels; a bilingual panel followed by a lay panel. First, a bilingual panel produces an initial translation of the questionnaire into the target language. The bilingual panel was conducted with presence of the original developer of ASQoL (SM), who advised on the closest meaning of the items and instructions to the original English language wording.”

Field testing

7) This section provides very little information on the actual procedure used here. What was the exact (standardized?) protocol of the CDIs? Was this based on established procedures? If so, please provide reference. Did the patients complete the questionnaires while thinking aloud? How were "difficulties and hesitation" observed and what (types of) questions were asked to examine comprehension and acceptability.

Reply: The CDIs were conducted by a trained and experienced researcher (YYL) according to standardized protocol to evaluate the face and content validity of the culturally adapted versions of ASQoL.

More information of the CDI is described below.
Cognitive debriefing interviews (CDIs) were conducted with Chinese speaking and English speaking patients with AxSpA to assess the relevance, acceptability and comprehensiveness of questionnaire items and instructions. AxSpA patients completed either the English or Chinese version of the ASQoL according to their primary language. Patients were chosen by the attending clinician to represent a range of disease severity, gender and age. Questionnaires were completed in the presence of a trained interviewer (YYL), who observed and made note of difficulties or hesitation in answering any items. The time of start and completion of the ASQoL were recorded. After completing the ASQoL, reasons for taking a longer time to complete specific items were elicited (e.g., whether the item wording was unclear, item was not relevant to the subject, etc). Patients were also asked whether all the items were relevant in assessing their QoL. With an open ended question, patients were invited to give further comments about the questionnaire and asked if any important aspects of their experience had been omitted. This method of evaluation of content validity fulfilled the recommended requirements, and has been used in the development of ASQoL (3) and other needs-based PROMs (10) (21).

Statistical analysis

8) Cronbach's alpha >0.70 is considered adequate for group-level use of the scores only.

Reply: We have specified this point in the article (Methods, page 11), and addressed as limitation in discussion.

Action: Discussion, page 23

“The internal consistency of the adapted ASQoL was 0.88 in this study, which is adequate for group-level use, and just marginally fall short of >0.90 criteria for individual level use. Therefore, the adapted ASQoL would be adequate for monitoring of QoL at least for group-level. ”

9) Were there no a-priori hypotheses with respect to the absolute and differential strength of the associations with other measures? This is highly desirable for adequate assessments of construct validity (see for instance Terwee et al) and just concluding that the scores are "moderately" correlated with a whole range of other patient-reported measures that are intended to measure different concepts is not very informative.

Reply: We have stated the expected magnitude of correlation for ASQoL with other PROMs. And the results aligned with our hypothesis that ASQoL had higher correlation with SF-36 subscales and summary scales; and less so with other PROMs that measure different construct. The construct validity of ASQoL was further supported by the known group validities and RASCH model analysis.
“Strong ($r \geq 0.7$) and moderate correlation coefficients ($r = 0.5-0.7$) suggest that the scores from two PROMs are measuring related construct whereas weak correlation coefficients ($r \leq 0.3$) suggest the PROMs are measuring different construct (34). We hypothesize the ASQoL should correlating with SF36 subscales at least moderately, and to a lesser extend with pain or global assessment indexes.”

Results, page 18:

“Moderate correlations were observed between ASQoL scores and SF-36 subdomain scores and summary scores, indicating that these impairments and functional limitations influence QoL. There were moderate correlations (but lesser extent) to pain, global assessments, and physical function. This aligned with the hypothesis testing of ASQoL as a measure of QoL.”

10) The fact that all other measures were patient-reported only (with the exception of PhGA) should be mentioned as an important limitation of this study in the discussion.

Reply: We have added this point in our limitation.

Action: Discussion, page 23

“The instruments for assessing patient’s status were all patient reported outcomes with the exception of PhGA, which may limit the comparison of ASQoL with a more objective assessment of disease activity. This is particularly true for the known-groups validity, which would be more robust if it was established with a more clinical and objective standard. Nonetheless, except Ankylosing Spondylitis Disease Activity Score (ASDAS) that takes into account acute phase reactants in blood, other assessments of AxSpA at the current moment heavily rely on PROMs (12). Besides, as the intended use of ASQoL is to measure the QoL in the patient’s perspective (45), it is relevant to use other PROMs for a comparison. It is also known that QoL is not necessarily strongly related to clinical severity.”

11) Using the same measures for both convergent validity assessment (correlations) and known-groups validity (dichotomized with t-test) essentially provides very little additional information. Again, it would have been very informative if know-groups could have been distinguished based on more clinical or objective standards.

Reply: This has been addressed together in question 10.

12) Spearman's correlation coefficient is not a suitable measure for test-retest reliability as this only assesses the association between rank orderings of patient scores and not agreement between scores (for instance if all patients have changed substantially, but in the same direction, this will still result in a high coefficient). Especially given the relatively small sample size of the test-retest sample (which is actually smaller than those recommended for instance by Terwee et al), the authors should report the intraclass correlation coefficient for agreement with 95% confidence intervals. Preferably supplemented with a Bland-Altman plot with limits of agreement. Together, these analyses would provide a much better picture of actual agreement between test-retest scores and systematic and unsystematic changes in scores.
Reply: We thank the reviewers comment. Although the use of ICC is common in reporting of test retest reliability, the use of ICC here is inappropriate. This is because the ASQoL is a parametric test. As the PsAQoL data is ordinal level, a non-parametric statistic such as Spearman's rank correlation coefficient should be used. We reported the ICC figure for information only.

Action: Results, page 20

“The intraclass correlation coefficient of ASQoL between time 1 and 2 was 0.86 (CI: 0.74-0.92).”

13) The authors should define cut-off points for both internal consistency and test-retest reliability for different purposes and interpret the findings accordingly in their conclusions. Generally, coefficients >0.70 are considered adequate for group-level use of an instruments, while coefficients >0.90 are considered necessary for individual level use (see also Terwee et al). Both the internal consistency (0.88) and test-retest reliability (0.81) do meet the latter cut-off.

Reply: We thank for reviewer for pointing out the deficit. We have made relevant comment in the discussion part. However, we have demonstrated evidence supporting the internal consistency and structural validity through the Rasch model analysis.

Action: Discussion, page 23

“The internal consistency of the adapted ASQoL falls marginally short of the recommended (>0.90) level for individual level use. However, the Rasch model analysis supported the necessary internal consistency and structure validity of the Singapore versions of ASQoL.”

Reply: As for the test-retest reliability, if we limited the analysis to patients who reported a change of BAS-G less than the reported MCID (15mm), the test-retest reliability was 0.85 which fulfilled the cut off criteria. However, the sample size for such analysis was small and we have addressed this in limitation.

Action: Discussion, page 23

“We acknowledge the slightly low Spearman’s coefficient (0.81) for test-retest reliability, and was improved to 0.85 when the analysis was limited to patients with more stringent criteria of “no change”. Of note is that the sample size for establishing this test-retest reliability (n=36) was considered rather low (32).”

14) Please describe how and where the sample for the test-retest study was recruited and was this sample similar to the cross-sectional sample with respect to demographic and clinical characteristics?

Reply: The cross sectional study and the test-retest study were separate studies, recruited patients from the same clinic at different period of time. As the primary question of the test-retest study was to address whether ASQoL was reliable while patient’s condition was stable, patients
recruited to this study was not consecutive, and not mean to represent the whole cohort of axial SpA. Patients having unstable condition and changing of medication/ exercise intervention were not recruited.

Action: Methods, page 15:

“We only included patients with stable SpA, and excluded those we were expecting a change in condition, such as those having changes in medication regimen or changing new exercise intervention or plan.”

Results

15) Field testing

Please provide more detailed information on actual comments and difficulties patients did have with the ASQoL instead of just stating that "overall patients found the instrument easy to understand and relevant".. and "…no changes were necessary".

Reply: We have included more information of the CDI in the results.

Action: Results, page 16-17

Cognitive Debriefing Interviews (CDI) were conducted with ten English speaking (age range 22-55 years, 90% male, 50% married, 20% unemployed) and ten Chinese speaking (age range 24 – 60 years, 60% male, 80% married, 20% unemployed) patients with AxSpA. Overall, patients found the instrument easy to understand and relevant to their condition. Five patients highlighted areas which affected their lives that may be missed. This included restricted neck and back mobility, lower body mobility, difficult to drive long distances, impairment in work productivity. After discussion, patients agreed the above areaed were assessed in item 1 (limits places can go), item 4 (struggle with chores), item 7 and 12 (tired), and item 8 (keep taking break when work). Therefore, there was no important aspects of the impact of their condition had been omitted.

Discussion

16) The authors conclude that findings are comparable to those found for the original UK measure, but this cannot be concluded from the current paper (e.g., not without describing these psychometric properties in the UK).

Reply: We have described the psychometric properties of the ASQoL in the Method part.

Action: Methods, page 10
“The original versions had excellent internal consistency (Cronbach’s alpha = 0.89-0.91), test-retest reliability (r=0.92 UK and 0.91 NL), and fulfilled the requirement of unidimensionality (fit to the Rasch model) (3).”

17) Please provide more thorough attention to the limitations and possible impactions of the current study and assumptions made. Some of these that should at the very least be described are the reliance on patient-reported measures only for validity, and the combining of English and Chinese version in the analyses.

Reply: We have addressed the limitation of reliance on patient-reported measures for comparison above in answer to question 10. We have also addressed the limitation of combining both language versions in the analysis in discussion.

Action: Discussion, page 19

“It is also a limitation to validate the combined Chinese and English versions of ASQoL for a practical consideration. However the language of administration (English or Chinese) did not affect the psychometric properties of the ASQoL, given the low level of DIF related to language, despite the small number of patients in the Chinese subgroup.”

18) The authors conclude that the scale can be used in routine practice (abstract), clinical trials and cohort studies (discussion), but it is not made clear which types of psychometric properties need to be established and which minimal standards a scale must meet, to be confidently used for these different types of applications. E.g., is a reliability of 0.81 sufficiently high for (individual?) use in clinical practice and can you really recommend the measure for cohort studies and RCTs without having examined responsiveness. Overall, I feel that recommending the scale on the current findings is overly optimistic and a more conservative claims and conclusions should be drawn given the considerable limitations, examined properties and outcomes of the current study.

Reply: We have aligned the abstract and discussion, and conclude that the adapted version can be used in clinical trials or cohort study settings. With the additional IRT analysis and explanation, I hope these response and revisions will be considered a satisfactory improvement to support our conclusion.