PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (http://bmjopen.bmj.com/site/about/resources/checklist.pdf) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

<table>
<thead>
<tr>
<th>TITLE (PROVISIONAL)</th>
<th>Assessment of Workplace Bullying: Reliability and Validity of an Arabic Version of the Negative Acts Questionnaire-Revised (NAQ-R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTHORS</td>
<td>Makarem, Nisrine; Tavitian-Elmadjian, Lucy; Brome, Dayana; Hamadeh, Ghassan; Einarsen, Ståle</td>
</tr>
</tbody>
</table>

VERSION 1 – REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Callegari Camilla</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Univ Insubria, Medicine and Surgery</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>23-May-2018</td>
</tr>
</tbody>
</table>

| GENERAL COMMENTS          | The article entitled "Assessment of Workplace Bullying: Reliability and Validity of an Arabic Version of the Negative Acts Questionnaire-Revised (NAQ-R)" by Makarem et al. is interesting and worthy of publication. The analysis and discussion of the psychometric properties of the scale seems well done and described. However, some points have to be revised, as suggested: 
- the exclusion criteria of participants should be better specified; 
- the distinction between strengths and limitations in the summary should be better specified, as well; 
- an explanation of the statistical tests performed and why authors have chosen the proper tests should be put in Methods section; 
- the abstract structure has to be in line with the manuscript structure (Introduction, in which aims of the study has to be specified; Materials and Methods, in which procedure, settings and participants have to be included; Results and Discussion). 
- the software used for the statistical analysis has to be specified in “Overview of statistical analyses” section; 
- dealing with the global evaluation done, depending on personal resources, it could have been useful an assessment of the resilience for example through the resilience scale (Reliability and validity of the Italian version of the 14-item resilience scale. Callegari C, Bertù L, Lucano M, Ielmini M, Braggio E, Vender S. Psychology Research and Behavior Management Open Access Volume 9, 3 October 2016, Pages 277-284); the lack of this evaluation should be defined as a limit and proposed as a future goal. 
In consideration of these points, the paper need minor modifications. |

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Senay KARADAG ARLI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agri Ibrahim Cecen University School of Health, TURKEY</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>08-Jun-2018</td>
</tr>
</tbody>
</table>
GENERAL COMMENTS

I think it is a good work.

REVIEWER

Tsair-Wei Chien
Chi Mei Medical Center, Taiwan

REVIEW RETURNED

14-Jun-2018

GENERAL COMMENTS

This article can be an interesting one that is worth researching in the field of applying Assessment of Workplace Bullying to verify the Reliability and Validity of an Arabic Version of the Negative Acts Questionnaire-Revised (NAQ-R). The premis of the study should be (1) a reliable and validate scale that can be used for assessing workplace bullying. (2) an useful sample obtained in this study. Finnally, explaining the survey results.

However, we criticize the structure of the study which lacks reasonable and logic thinking process as we mentioned above. That is, authors reported the results of the study and analyzed the scale quality by reliability and validation, particularly the scale presenting unacceptable in reliability and validation.

We illustrate several comments as below:

Major concerns:

1. The article can be concise more to let the paper readable and comprehensive, for instance, removing many sections irrelevant to the reliability and validation of NAQ-R, such as the last paragraph of the present study or the pre-paragraph prior to the section of Subjects and Methods.

2. We are glad to see the logic layout before the section on Subjects and Methods and expect to see the purposes of this study. However, none of these purposes were illustrated there.

3. We see the three model regarding one, two and three factors were examined as not complying with the acceptable level of validation using CFA on page 18 and 19. We confused why and what are those EFA and CFA in pages 21-22 again performed in this study.

4. We cannot merely see the reliability because the reliability is dependent on the item length. The result regarding CFA on page 18 and 19 was not acceptable. All the ANOVA or others in this study is in vain. If the instrument consists of several factors or subscales, the overall reliability(0.90_in Table 4 is not necessary because two subscales with two distinct domains cannot be combined to assess any without meaningfulness.

5. Due to the reason of the scale without validation, we would not like to see any anymore for the comparison or others.

6. In the literature, the NAQ-R should be a single dimensional scale, see the one at https://www.ncbi.nlm.nih.gov/pubmed/24534113. If any not complying with the conditional, the summation score, for instance in the page 16 is unnecessary.

Minor concerns:

1. In Table 1 and 2, I suggest authors to separate the gender into two or three(unknown) columns to let the cross tables presenting more information for readers.

2. In Discussion, I expect to remove any irrelevant part of NAQ-R in reliability or validation and to involve more structure sections in this study, such as 1) What This Knowledge Adds to What We Already Knew;(2)What the Findings Imply and What Should Be Changed?(3) Strengths of This Study; and (4)Limitations and Future study besides (5) conclusions.

3. In Abstract, Authors stated that A 14-item two-factor NAQ-R, with subscales of person and work-related bullying was supported. Reliability coefficients for total and subscale scores of the NAQ-R ranged from 0.63 to 0.90 in Table 4. Readers will be confused with why the three subscales existing in Table 4.

In all, I conclude that any one instrument without validation results in any follow-up analysis problematic and unnecessary. In this study, we cannot see any supporting the NAQ-R fitting to the validation model. That is, the reason might be person responding to the NAQ-R with misbehaviour or careless or guessing and resulting in the instrument unacceptable.
Response to Reviewer 1: Thank you for your review.
- Exclusion criteria clearly delineated on page 11 under the subtitle "Subjects and methods"
- Distinction between strengths and limitations in the summary is better specified.
- Statistical tests and why they have been chosen have been deleted from the results section (see track changes under results sections) and have been added to the methods section under the subheading "overview of statistical analyses" on pages 14-15.
- The abstract structure has been changed accordingly.
- The software used for statistical analysis has been specified in the last sentence of "Overview of statistical analyses" on page 15.
- The concept of resilience and its assessment scale has been added in the discussion section on pages 29-30. The limitations section was updated to include the lack of this evaluation as a limit and proposed as a future goal.

Reviewer 2: Thank you for your review.

Reviewer 3: Thank you for the review.

Major concerns:
1. Indicated paragraph has been deleted to make the study more concise.
2. The logical layout explaining the purpose of this study has been added to the second paragraph under the section "The present Study" on pages 9-10.
3. The rationale behind using the EFA and CFA in this study is for the following reason: the three tested models were not good fits for the data; hence, cross validation (Bollen, 1989) was implemented so as to assess the structure of the Arabic NAQ-R.
After randomly splitting the data into two roughly equal halves, we applied exploratory factor analysis (EFA) extracting factors with eigenvalues greater than 1.0 with oblique rotation on the first half and then applied a CFA on the second half using the factor structure emerging from the EFA.
4. Please provide further explanation, point is not clear to us.
5. Please provide further clarification as idea not understood.
6. Although the NAQ-R may be used as a single factor measure, Einarsen, Hoel, and Notelaers (2009) showed a three-factor solution for the NAQ-R: personal bullying, work-related bullying, and physically intimidating forms of bullying. Additionally, confirmatory factor analysis (CFA) implemented in a study conducted by Gupta, Bakhshi and Einarsen (2017) also revealed a best fit three-factor model of work-related, person-related and physically intimidating bullying for NAQ-R. The three factor structure was also supported by Tsuno, Kawkamu, Inoue and Abe (2010).

Minor concerns:
1. Since we found no differences on the NAQ-R total and subscale scores between males and females, we did not separate gender into columns in the tables.
2. Discussion section has been modified according to the remark given.
3. In the Abstract, the final finding was reported whereas the reliability subscale scores in table 4 report results findings that were concluded from the statistical analysis. We believe that the NAQ-R fitting to the validation model was not supported not mainly due to the misbehavior or careless guessing of the participants, but to other reasons as well. A significant reason can be attributed to the cultural variations in perception of nature of bullying especially that NAQ-R was developed in the west but the study was performed in an Arab culture.