Shiraz E-Medical Journal
Volume 22 · Issue 12 · e110977

OPEN PEER REVIEW

A Differential Item Functioning (DIF) Study of the Infection Control Standard Isolation Precaution Instrument Across Gender and Major Among Healthcare Workers

Amin Mousavi, Mohsen Momeni, Mina Danaei, Mehrdad Askarian

SUBMITTED	REVISED	ACCEPTED	DOI
7 Nov 2020	9 Jun 2021	17 Jun 2021	10.5812/semj.110977

About this report

This open peer review (OPR) report shows the public comments made on this article during peer review, by the reviewers, the associate editor and the editor-in-chief, for every revision round, together with the authors' replies. Comments appear only with the consent of the people who wrote them. Reviewers who chose to sign their report are named; the others are shown as "Referee". Comments are edited for language only, and confidential remarks to the editors are not included.

More about open peer review: brieflands.com/knowledgebase/opr

Contents

1 Original submission 2
2 Revision 1 Associate Editor 3
Authors' reply to reviewers 4
Reviewer and editor comments 7
3 Revision 2 Associate Editor 8
Authors' reply to reviewers 9
Reviewer and editor comments 12

1 Original submission

0 comments

No public comments in this round.

2

Revision 1

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Response to reviewers

We would like to thank anonymous reviewers for their constructive comments. We carefully reviewed all of them and incorporated their suggestion. Where needed, we added extra information to the manuscript in order to increase the paper's readability and clarity. All changes/additions are highlighted with Red font in the text. Below you can find our detailed reposes to each points made by the reviewers:

Reviewer 1:

1. The title needs revision. Title should indicate the objective of the study. For instance, "measurement invariance of the standard infection control isolation precaution instrument across gender and major-neutrality" or "differential item functioning of the …"

Following your recommendation, we have changed the title to "**A Differential Item Functioning (DIF) study of the standard infection control isolation precaution instrument across gender and major among healthcare workers"**

2. Non-differentia item functioning" is not a common term. This sentence should be revised.

Thanks for your note, we changed it to "the absence of Differential Item Functioning"

3. At the end of introduction, the consequence of lack of measurement invariance should be stated more clearly.

The following paragraph has been added for more clarity "Measurement invariance evaluates the equivalency of a construct across groups or measurement occasions. Measurement non-invariance implies that the construct under study is not comparable across groups or occasions. This means that, under lack of measurement invariance, simply using observed scores for statistical analysis such as ANOVA would yield incorrect conclusions because the scores have already different meanings across groups. Therefore, doing an ANOVA would be analyzing as set of data that is already different. This emphasizes the importance of establishing measurement invariance before using observed score for any statistical analysis."

4. In the "Participant and data" section it would be better to summarize the mean of age and scores in a table or report Mean and standard devotion as Mean±SD.

Thanks for your suggestion, we chose to report them as Mean±SD.

5. Average should be substituted with Mean.

All the instances have been changed to mean.

6. Differential test functioning should be explained more clearly.

We added the following description "Differential Test Functioning compares a set of items across groups. In other words, it evaluates the comparability of total test scores (i.e., observed scores) across groups. In this sense, it is an extension of DIF" to the Data analysis section.

7. The Mantel-Haenzel test is used for binary response questionnaires. However, it was explained that the knowledge subscale has three possible answers and practice subscale has 5 possible answers. How the authors apply this method for these kinds of items.

The original formula for Mantel-Haenzel was develop for K 2×2 contingency tables in 1959 and its extension to K I×J contingency tables was introduced in 1963 by Mantel. As it is mentioned under section 3.2. "A score of 1 point was allocated to 'always' and 0 for all other answers." for the practice sub-scale converting it into a dichotomous item.

8. In the second paragraph of the Results section, it was written that "a review of figures for …."; but there is no figure in the manuscript.

In this text, the word "figure" means "number". We replaced all instances with "values" for more clarity.

9. The abbreviations used in Tables 1 and 2 should be introduced in the caption of the tables. It is not clear what the MH CHI,… or ETS,…A, B Or C used in the table.

Thanks for this point. All the abbreviations were introduced in the Data analysis section but we added them as notes to the Table for more clarity. Additionally, we added * to show items with statistically significant MH Chi-square values and used **bold** font to indicate DIF items based on LOR Z in both tables.

10. There is no cut-point for the values of each criterion used in the Table. This was should be explained in the Methods section.

The cut-points were introduced in the data analysis sections and we added more informative details heling readers understanding the criteria for flagging an item as DIF . We also added the classification criteria for ETS categories as follows:

These categories are based on differences calculated on the scale of item difficulty, known as Mantel-Haenszel delta difference (MH D-DIF). A statistically non-significant MH D-DIF or less than 1.0 in absolute value would categorize an item as A (i.e., small size DIF). Items with MH

D-DIF significantly greater than 1.0 and absolute value of 1.5 or greater would be categorized as C (i.e., large DIF). All other items would be categorized as B (i.e., medium size DIF).

11. How we can find that all items of the practice subscale function similarly across gender". This issue should be explained in the Methods section; based on which criterion we can conclude that which item function similarly or differently across the groups.

Thanks for your comment. We initially assumed that readers would be able to link tables with criteria outlined in the Data analysis section and conclude identification of DIF. Following your recommendation, additional details are now added to the text pointing out to significant MH chi-square and magnitude of DIF for corresponding items so readers would know the reasons behind each item being flagged as DIF.

12. There is no comparison with similar studies in the Discussion section. Only the results were repeated and the possible reasons for the findings were provided without any reference to documented articles. This section should be rewrite substantially.

This is the first DIF study done for this instrument and there is no other similar studies. Future DIF studies can refer to this study and compare their findings.

2 Revision 1

1 comment · 3 Jun 2021

Associate Editor

MINOR REVISION 3 Jun 2021

Dear Authors
Thank you for revising your manuscript. However, according to one of the referees, you have to deal with three issues. If you are interested in continuing the process, please answer the reviewer's comments one by one and resend the revised version of your manuscript.
Kindest regards
Editorial team

3

Revision 2

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Response to reviewer 1's comments

We would like to thank the reviewer1 for his/her constructive comments. We added extra information on DIF and consequences of measurement non-invariance. We made all changes/additions considering the word limits of the journal. The changes/additions are highlighted with Green font in the text. We believe that the added explanations alongside with the provided references allow interested readers gaining more in-depth knowledge related to this manuscript. Below you can find our detailed reposes to each points made by the reviewers:

Reviewer 1:

1-First, the consequence of lack of measurement invariance is not well written. When measurement invariance does not hold the mean cross-group comparison is not valid since it is not clear whether the observed difference between the groups is a real difference or due to different perceptions of individuals

We believe that we exactly talked about this point but in different way. When we referred to ANOVA and the effect of measurement non-invariance on its results, we were actually talking about group comparisons and how they can be affected by the lack of measurement invariance. Following Reviewer 1's comment, we adopted a simpler and clearer language. We also talked about the effect of measurement non-invariance on repeated measures designs and when assessing effectiveness of interventions. This is what readers would see in the text:

Measurement invariance (MI) evaluates the equivalency of a construct across groups or measurement occasions. MI is relevant to group comparisons (e.g., ANOVA), comparing means across repeated measures (e.g., pretes-posttest designs) and comparing relationship between constructs across groups. Measurement non-invariance, on the other hand, implies that the construct of interest is not comparable across groups or occasions. This means that, under lack of measurement invariance, simply using observed scores for statistical analysis such as ANOVA would yield incorrect conclusions because is not clear whether the observed difference between the groups is a real difference or due to different perceptions of individuals. One other occasion that measurement non-invariances can generate incorrect conclusions is when assessing the effectiveness of interventions. Such studies usually involve pretest-posttest measurements, or control-treatment groups or both. Researchers need to ensure that scores from pretest and posttest or scores from treatment and control groups yield similar meanings and psychometric properties. If scores from pretest and posttest or scores from treatment and control groups are noninvariant then it would not be clear if an observed change was due to the intervention or a change in participant's perceptions of the construct under study. This emphasizes the importance of establishing measurement invariance before using observed score for any statistical analysis. (16).

2-Second, the explanation of DIF added to the manuscript is very short and it is not clear and should be explained more clearly with details..

Our main goal in this manuscript is to avoid detailed technical-statistical terms and as our main audience is practitioners. Following Reviewer 1's comment we extended the definition of DIF and illustrative figures. This is what readers would see in the text:

In technical terms, it is important to test whether the probability of responding to a specific item differs across different identifiable groups after controlling the construct being measured (13). This is known as testing for Differential Item Functioning (DIF). Generally speaking, there are two different types of DIF. The uniform DIF which implies constant differences of the probability of responding to an item between groups along the continuum of respondent's knowledge or ability (14). Figure 1 shows an example of uniform DIF.

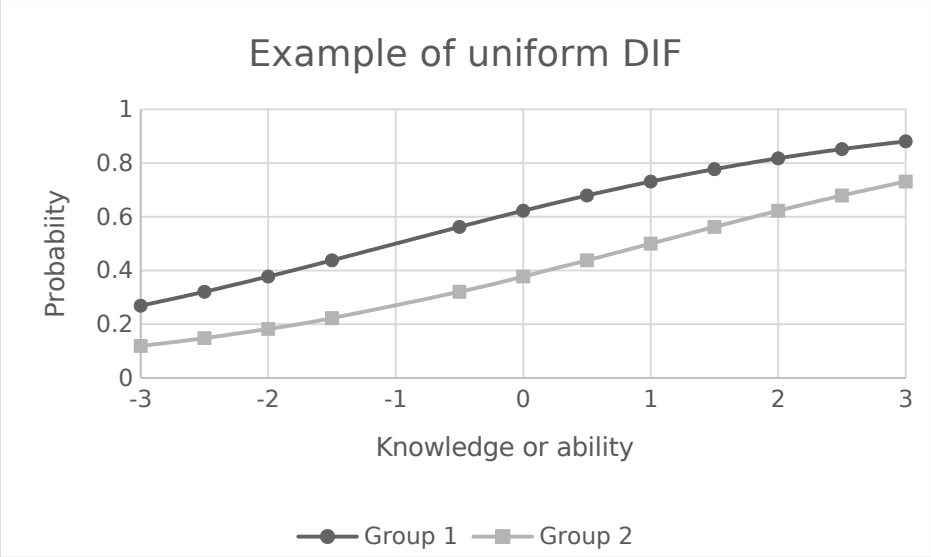


Figure 1: example of uniform DIF across two groups

The other type of DIF is known as nonuniform or crossing DIF. A nonuniform DIF implies that differences in the probability of responding to an item not only depends on respondent's group membership but also depends on his/her knowledge or ability. In another, there's an interaction between group membership and knowledge or ability. Figure 2 shows an example of nonuniform DIF.

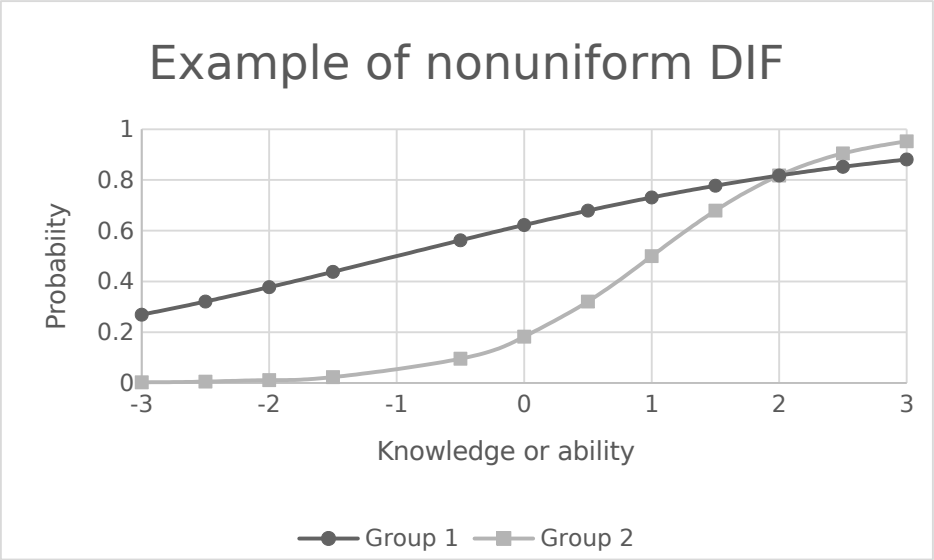


Figure 2: example of nonuniform DIF across two groups

The best case scenario is not observing DIF. This means that the probability of correctly answering a question should only be dependent on the respondent's knowledge or ability and no other external variable such as gender, social status, job status or race

3 Revision 2

1 comment · 13 Jun 2021

Associate Editor

ACCEPT 13 Jun 2021

Dear Authors
We are pleased to inform you that your manuscript entitled "A Differential Item Functioning (DIF) study of the standard infection control isolation precaution instrument across gender and major among healthcare workers" has been primarily accepted for publication in SeMJ. Please follow the guideline for the rest of the process.
Kindest regards
Editorial Team