

Differential item functioning (DIF) and differential test functioning (DTF) analysis of the Urdu Four-Dimensional Questionnaire (4DSQ)

Berend Terluin, Farzana Ashraf

Introduction

The Four-Dimensional Questionnaire (4DSQ) is a self-report questionnaire measuring distress, depression, anxiety and somatization [1]. The questionnaire has been developed in Dutch primary care and was translated into the Urdu language. This report concerns the psychometric validation of the Urdu version of the 4DSQ against the original Dutch questionnaire.

Methods

Translation

The first Urdu version of the 4DSQ was translated in 2013 by a Pakistani interpreter, using the English version of the 4DSQ. Subsequently, the Urdu 4DSQ was examined and compared to the English version by two bilingual experts and psychologists. On the basis of their reports, a few changes were made to obtain better and more genuine Urdu responses.

Data for psychometric analyses

Urdu 4DSQ data were collected in 205 non-clinical subjects. The sample was recruited conveniently from a university (teachers) and a commercial bank (employees) and through personal reference (particularly house wives). Included were 50 male and 155 female adults between 30 to 45 years of age. The Dutch 4DSQ reference data were obtained by drawing a gender and age matched sample ($n = 308$) from a database of questionnaire data of patients with (probable) mental health problems collected in a primary care health center in Almere, the Netherlands. Three reference patients were selected for two Urdu subject. Missing item scores were imputed using the response function method [2].

Analysis

Whether the Urdu 4DSQ measures the same constructs in the same way as the Dutch 4DSQ, was examined using differential item functioning (DIF) analysis and differential test functioning (DTF) analysis.

The idea behind DIF-analysis as a way to validate questionnaire translations, is that the translated scale measures the same as the original scale when the translated items can be shown to “function” the same as the original items [3]. For the detection of DIF we used the hybrid ordinal logistic regression method as implemented in the R-package “lordif” [4]. The criterion was an R-square change of >0.02 when including group membership and its interaction with the scale score in the regression model.

In order to assess the impact of item-level DIF on the scale scores (i.e., DTF), we calculated Rasch theta-scores adjusted for DIF, using concurrent calibration. Next, ordinary scale (sum) scores were plotted as a function of the DIF-free theta-score and group membership. The DTF becomes apparent in the vertical distance between the curves.

Results

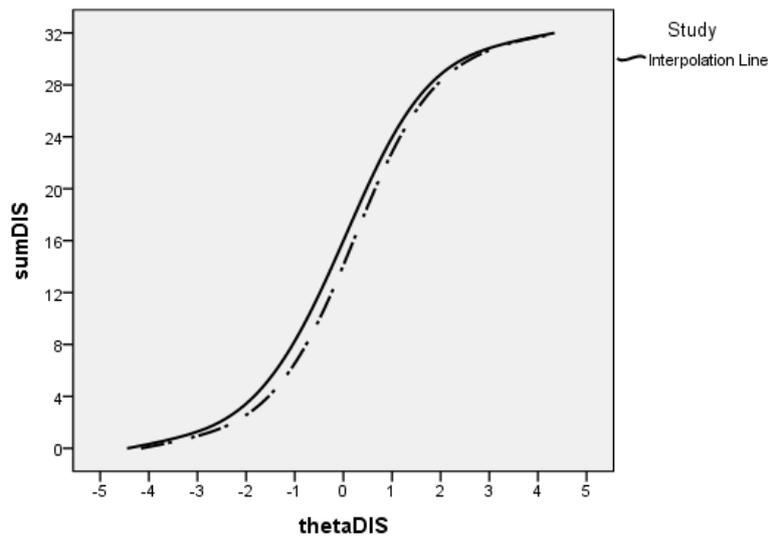
Significant DIF was detected in 17 items (Table 1). Most DIF-laden items were more severe for Urdu speaking people, meaning that Urdu people need more severe levels of distress, depression or anxiety to endorse these items. Logically, items that are more severe yield lower item and scale scores. Some items were less severe, meaning that Urdu people needed less severe levels of distress, depression or anxiety to endorse these items compared to Dutch people.

Table 1: 4DSQ-items detected as having significant DIF*

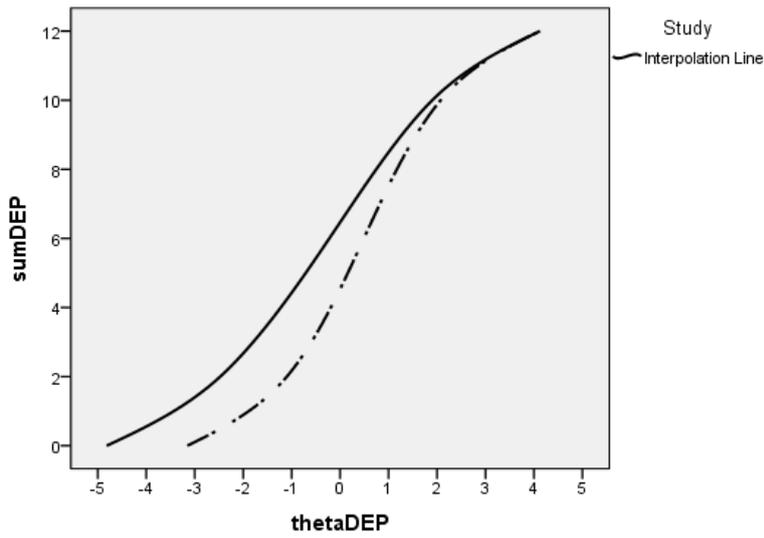
4DSQ-scale	Item numbers
Distress	19, 20, 25, 26, 39, 48
Depression	28, 34, 35, 46
Anxiety	21, 27, 40, 43, 49, 50
Somatization	3

* **red**: the item was more severe for Urdu people compared to Dutch patients
green: the item was less severe for Urdu people compared to Dutch patients
black: the item demonstrated "non-uniform" DIF

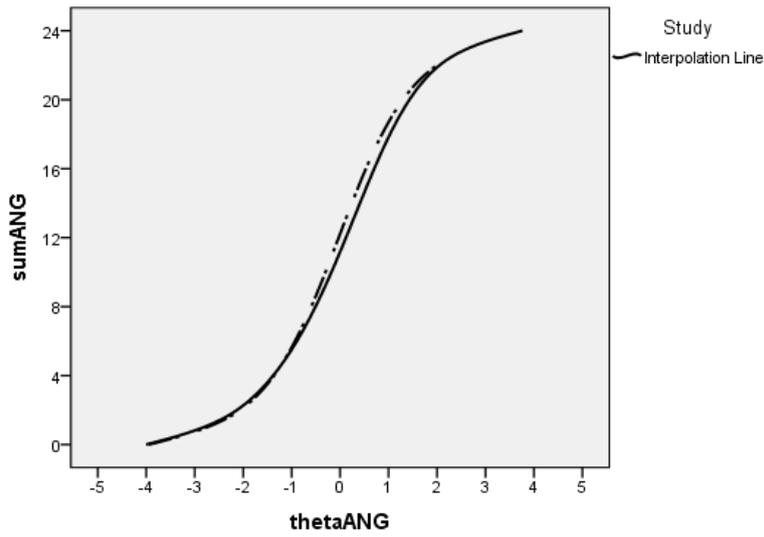
DTF-analysis demonstrated that there was some DTF in the distress scale and substantial DTF in the depression scale. The vertical distance between the curves is about 1.5-2 points in case of the distress scale, and 2 points in case of the depression scale. The anxiety and somatization scales did not show much DTF. In the anxiety scale this may be due to "DIF cancellation", that is the effect of more severe items is cancelled out by the opposite effect of less severe items.



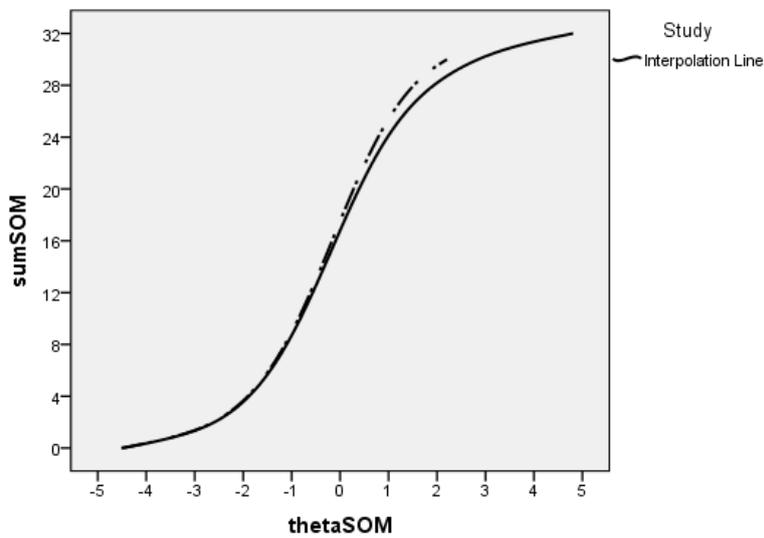
Distress scale



Depression scale



Anxiety scale



Somatization scale

Figure 1: Differential test functioning of the Urdu 4DSQ compared to the original Dutch questionnaire (dotted line: Urdu, solid line: Dutch).

Discussion

The good news is that the Urdu translated 4DSQ scales measure the same constructs as the original Dutch scales do. For example, when someone scores “moderate” on the Urdu distress scale, it means that the person has moderate distress, the same kind of distress as experienced by someone who scores “moderate” on the Dutch distress scale. However, the bad news is that, because some translated items “function” a little differently from their original Dutch counterparts, the Urdu 4DSQ does not measure the same constructs in exactly the same way. For instance, a distress score of 13 on the Urdu distress scale corresponds to a distress score of 15 on the Dutch distress scale. Urdu people score a little lower than expected on the distress and depression scales, and they tend to score a little higher on the anxiety and somatization scales.

Most DIF-laden distress items are more severe for Urdu people. That is, the translated items in Urdu represent a more severe level of distress than the original Dutch items do. Urdu people need, therefore, higher levels of distress to endorse these items, compared to Dutch people. Consequently, Urdu people obtain 1.5-2 points less on the distress scale than Dutch people when having the same level of true distress.

The same is true for the depression items. Most DIF-laden items are more severe for Urdu people and consequently they score about 2 points lower than Dutch people with the same levels of depression. The problem of the depression scale is more serious than the distress scale because the depression scale ranges only 12 points. A 2-point difference represents thus 17% of the scale range.

With respect to the anxiety and somatization scales, we can see that Urdu people tend to score a little higher than Dutch people would do, but only when anxiety or somatization is raised to relatively high levels.

What should we do? Now that we know that the Urdu version of the 4DSQ measures distress, depression, anxiety and somatization much as the original Dutch questionnaire does, the Urdu 4DSQ can be used for that purpose. However, in interpreting the scores DIF must be taken into account. The easiest way to do so, is to adjust the cut-off points of the Urdu 4DSQ scales (Table 2). That way, using the cut-off scores as bench marks, the Urdu scores can be interpreted correctly.

In the future, some of the DIF-laden items could possibly be replaced by better items, so that adjustment of cut-offs is no longer needed. New items, however, should be tested before they can be included in an Urdu translation 2.0.

Table 2: Proposed adjusted cut-off points

Scale	Interpretation	Dutch scores	Urdu scores
Distress	low/normal	0-10	0-8
	moderately elevated	11-20	9-19
	strongly elevated	21-32	20-32
Depression	low/normal	0-2	0
	moderately elevated	3-5	1-3
	strongly elevated	6-12	4-12
Anxiety	low/normal	0-3	0-3
	moderately elevated	4-9	4-9
	strongly elevated	10-24	10-24
Somatization	low/normal	0-10	0-10
	moderately elevated	11-20	11-20
	strongly elevated	21-32	21-32

Reference List

1. Terluin B, Van Marwijk HWJ, Adèr HJ, De Vet HCW, Penninx BWJH, Hermens MLM *et al.*: **The Four-Dimensional Symptom Questionnaire (4DSQ): a validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization.** *BMC Psychiatry* 2006, **6**: 34.
2. van Ginkel JR, van der Ark LA: **SPSS syntax for missing value imputation in test and questionnaire data.** *Applied Psychological Measurement* 2005, **29**: 152-153.
3. Petersen MA, Groenvold M, Bjorner JB, Aaronson N, Conroy T, Cull A *et al.*: **Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire.** *Qual Life Res* 2003, **12**: 373-385.
4. Choi SW, Gibbons LE, Crane PK: **lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations.** *J Stat Softw* 2011, **39**: 1-30.