Reviewer’s report

Title: The 12-item World Health Organization Disability Assessment Schedule II (WHO-DAS II): A nonparametric item response analysis

Version: 2 Date: 6 April 2010

Reviewer: Russell Steele

Reviewer’s report:

In "The 12-item World Health Organization Disability Assessment Schedule II (WHO-DAS II): A nonparametric item response analysis ", the authors present an assessment of the psychometric properties of the WHODAS-12 using an IRT modeling approach. They have a reasonably large cohort of patients (3615 general practice patients with a major depressive episode) and they use somewhat advanced methods in order to analyze the data. They conclude that the items for the WHODAS-12 are sufficiently discriminatory and are useful for the purpose of measuring patient disability. They also find only modest evidence of Differential Item Functioning (DIF) in their cohort with respect to gender.

I thought this was an interesting paper and their statistical methods (my expertise area) are quite sound. I think that the WHODAS-12 is quite an underutilized questionnaire in the HRQoL literature (particularly relative to the SF-36), so it is important for these types of papers to validate its use in studies and increase its exposure (and the research community's general trust in its use). The conclusions are clearly supported by the results and so, in my opinion, this is a correct piece of research (and certainly publishable on those grounds).

No major compulsory or minor essential revisions.

Discretionary Revisions:

As for my limited criticisms of the paper, they would be merely of style rather than of substance. In general, I prefer parametric methods combined with assessment of parametric assumptions rather than non-parametric methods mostly due to power arguments. In a perfect world, one would use both methods (parametric and non-parametric) and they would agree on the general interpretation of the available evidence, giving lots of strength to the results. Here the authors only present the non-parametric results (and somewhat weak reasoning for doing so). It would have been nice to see the parametric IRT results (or at least for the authors to have tried them and mention a brief comparison in the text). In particular, the null DIF results could be due to lack of power due to the use of the non-parametric smoothing technique (in place of a parametric IRT curve). Given the modest evidence for DIF here, it is difficult to conclude one way or the other. A parametric IRT analysis would have potentially had more power and given a different (albeit not necessarily correct) answer. I had wished that the authors would have presented that result in their work here.
The only other complaint that I had about the paper is that they seem to under-interpret the results with respect to the items. They merely report parameter estimates and summaries of the various discrimination and difficulty ratings for the items. I would have liked to see more about the specific estimates for the difficulties and discrimination in relating the items to each other. For example, the ICC and OCC curves by item seem somewhat similar in shape. What does that say about the items themselves? Are they all of the same difficulty or are they measuring the same thing? Is there some way with the non-parametric approach to summarize differences amongst items in the way that one could with a 2 or 3-parameter IRT model? I want to stress that I'm not saying that non-parametric method they've used here is not the correct one to use; rather, that it would have made for a stronger and more interesting paper with more interpretation comparing curves between items. Similarly, the options comparisons seem important in that many of the plots are similar in Figure 4, but can we thusly conclude that the items all have similar discriminatory patterns with respect to the options?

In conclusion, I think this is a nice paper that really only needs to delve a bit more deeply into interpretations of their advanced (and justifiably so) statistical analyses in order to become a very good paper.

Russell Steele
Department of Mathematics and Statistics, McGill University

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests.