Reviewer’s report

Title: A Rasch analysis of the Manchester Foot Pain and Disability Index

Version: 2 Date: 12 October 2009

Reviewer: Gabrielle van der Velde

Reviewer’s report:

(Comments numbered 1.d to 11. refer to the numbering used in the Authors’ responses)

The reviewer thanks the authors for their responses to review comments. Below are issues which need to be addressed further.

1.d. The FPDI was designed to measure the construct ‘foot disability’. In the Background section, the authors describe the FPDI subscales as each measuring a single construct, for a total of four constructs. These are more correctly described as subscales that measure factors that map onto the construct foot disability.

2. The summative score is not ‘at best, an ordinal score’ if the FPDI consists of 4 constructs as the authors describe in the Background section (see comment 1.d) above. Even for ordinal-level scaling, a scale must be unidimensional to validly summate response data into a single score. The authors should correct this sentence.

11. 36 observations are an insufficient number of observations to conduct a Rasch analysis of the appearance subscale. The reviewer is not familiar with any sample size recommendation that supports such a small sample size, even Linacre. The reviewer suggests that the analysis of the appearance subscale should not be included due to the small sample size and fact that there are only 2 items which preclude selected analyses.

Miscellaneous:

Discussion:

1. Page 10, incomplete sentence (see below): what should it read? The reviewer also disagrees with the interpretation of this result: it suggests a lack of unidimensionality, not ‘could indicate’ as the authors have written.

‘There was some evidence of differential item functioning (DIF) by age on the item relating to avoiding rough and hard surfaces on the function scale, which could indicate a lack of unidimensionality in the ?????[31].’

2. There were insufficient observations for conducting a Rasch analysis of the appearance subscale (n = 36); the authors state that they have cited Linacre, but
they have not achieved his suggested requirements for adequate sample size for this subscale. Their argument that the p-value supports the adequate size of the sample is not relevant. The issue is whether their sample size was sufficient to observed differences between the response data and that expected by the Rasch model, not the statistical significance of the resulting statistic.

3. In the Discussion the authors state that they used the Rasch unidimensional measurement model to produce interval-level scores – but where are these scores? There are no Rasch-weighted scores for each of the subscales that are provided for users of the FPDI unless I missed them?. The reviewer suggests that the authors simply conducted a Rasch analysis to determine whether the response fit the Rasch model, which, if they do, demonstrates that the response data achieved interval level measurement. The authors interpret their results to mean that the FPDI subscales’ response data did fit the Rasch model, and therefore, do represent interval-level scaling. Certainly the authors should not include the Appearance subscale in this interpretation or conclusion, for the reasons discussed above, including the inordinately small sample size and the 2-item subscale.

Overall:
1) The authors should not present the results of analysis conducted on n = 36 observations for the appearance subscale as this sample size is insufficient; at the very least, the authors should not make any conclusion that suggests this subscale represents unidimensional or interval level scaling and the results should be presented with strong reservations and comments about the serious limitations which preclude meaningful results.

2) The conclusions regarding the interval-level scaling of the remaining 2 subscales are too strongly worded, with too much deference to results that support the interval scaling and too little consideration of the results that do not support their interval scaling. See for example the issue of unidimensionality w.r.t. the function subscale. This study is an initial exploration into the scaling properties of the 3 subscales in a small sample of persons; there was no interval level scoring provided (unless I missed it?) since there were no changes made to the subscales (e.g. item deletion, collapsing of response options) and since there is no exchange table provided between simple summed scores and Rasch-weighted scores. There are results that suggest that the two subscales do achieve interval level scaling, and also results that they do not. Further analyses are required to determine if these results are reproducible using larger sample sizes.

Quality of written English: Acceptable

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