

**Author's response to reviews**

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

**Authors:**

Sara Muller ([s.muller@cphc.keele.ac.uk](mailto:s.muller@cphc.keele.ac.uk))

Edward Roddy ([e.rodny@cphc.keele.ac.uk](mailto:e.rodny@cphc.keele.ac.uk))

**Version:** 3 **Date:** 19 October 2009

**Author's response to reviews:** see over

Arthritis Research Campaign National Primary Care Centre  
Primary Care Sciences  
Keele University  
Keele  
Staffordshire  
ST5 5BG

19 October 2009

Dear Profs Menz and Potter,

We are pleased to resubmit the following article, now entitled "A Rasch Analysis of the Manchester Foot Pain and Disability Index" for submission to the Journal of Foot and Ankle Research as a research article.

We thank the reviewers for their further comments and have addressed each of the points raised below.

Yours truly,  
Sara Muller

Mrs Sara Muller  
Arthritis Research Campaign National Primary Care Centre  
Primary Care Sciences  
Keele University  
Keele  
Staffordshire  
UK  
ST5 5BG

Email: [s.muller@cphc.keele.ac.uk](mailto:s.muller@cphc.keele.ac.uk)

Tel: +44 (0)1782 734853

Fax: +44 (0)1782 733911

## Response to reviewer's comments

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.

**We agree with the reviewer that FPDI constructs are more accurately described as subscales than constructs and have ensured that the correct term ("subscale" or "factor") is used throughout the manuscript.**

**However, we do not agree that the FPDI was designed to measure the construct "foot disability". The FPDI was designed to measure foot pain and disability. As we describe in our description of the development of the FPDI, inserted at the request of this reviewer, the included items were developed from a series of patient interviews that covered a number of issues including pain, activity limitation and footwear in addition to disability. We feel therefore that it would be incorrect to describe these subscales as mapping 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.

**We apologise that the reviewer did not think that our explanation of this point was clear enough. We intended 'at best' to mean, should the scale be unidimensional, it could only be ordinal. We have now amended this section of the text as follows:**

**"However, these summated totals were not suitable to correctly examine changes in score over time, or differences in scores between groups, because they were not shown to be unidimensional and were not of an interval-level, i.e. where a difference of, say, two points on the score is equivalent at all points along the continuum."**

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.

**We have stated at the beginning of the Results section under 'Study sample' that there are too few people available for the analysis of the appearance subscale in this dataset and have therefore not analysed this subscale:**

**"Hence, having removed those with extreme scores, 131 people were available for the derivation of the function subscale score, 133 for the pain subscale and 36 for the appearance subscale. This sample size for the appearance subscale was considered to be too small to allow assessment of the subscale's properties, and so further analyses of the two appearance items were not undertaken."**

**We have removed all further reference to the appearance subscale in the Results section and suggested in the Discussion that others may like to consider this subscale, although the small number of items makes this problematic:**

**"These analyses have shown that the function and pain subscales of the FPDI are unidimensional and that interval level scores can be obtained from the items of these subscales. It was not possible to assess the measurement properties of the appearance subscale due to the small number of people without extreme responses on this subscale. This is perhaps not surprising, as the appearance subscale consists of only two items, making scoring problematic."**

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].'

**We thank the reviewer for pointing out this typographical error and have inserted the words “in this subscale” at the end of the sentence. We have not however changed the wording of this sentence, as this is an empirical testing process and there is not evidence from either the t-tests or residual correlations to indicate a lack of unidimensionality in this subscale and removal / splitting of the item for DIF did not change the model. The most likely explanation seems to be a Type I statistical error. We have added this comment to the manuscript.**

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.

**As we have stated above in response to Comment 11, we have removed the appearance subscale from the analyses.**

**In relation to our comment on p-values, we are not claiming to have sufficient statistical power simply because we have reasonable p-values, just stating that substantially larger sample sizes would be needed in order to change the conclusions of the study about the fit of the data to the Rasch model (the item-trait interaction test) and that the sample size was large enough to result in what we suspect was a Type I error.**

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 FPD I 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 FPD I 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.

**We have discussed the methods of obtaining an interval level score from the FPD I item extensively in the Discussion section, where we have compared the potential use of a Rasch analysis in each dataset where a score is required to a conversion table. We have not included such a table because, as we have concluded, the most appropriate method of obtaining the score is dependent on the context, and we would not wish to encourage the inappropriate use of such a table.**

**We have not concluded that interval-level scaling is possible from the appearance subscale, and as we have responded above, this subscale has now been removed from the analyses.**

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.

**We did not suggest that interval-level measurement was possible from the appearance subscale, but this analysis has now been removed from the paper.**

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.

**We disagree with some of these comments, as detailed and explained above. We have made suggested changes in response to others (also detailed above).**