**Measuring burnout in paediatric oncology staff. Should we be using the Maslach Burnout Inventory?**

**Authors**

Dr Suzanne Mukherjee, Bsc. (Hons), Msc., PhD.

Research Fellow, Social Policy Research Unit, University of York, York, Y010 5DD.

Email: [suzanne.mukherjee@york.ac.uk](mailto:suzanne.mukherjee@york.ac.uk); Tel:01904 321950; Fax: 01904 321953

Professor Alan Tennant, BA, PhD., MD (honoris causa)

Permanent Visiting Professor, Faculty of Humanities and Social Sciences, Luzern University, Luceren, Switzerland. 

Professor Bryony Beresford, Bsc. (Hons.), PhD.

Research Director, Social Policy Research Unit, University of York, York.

**Correspondence to principal author:** Dr Suzanne Mukherjee

**Abstract**

Burnout in health service staff is a cause for concern since it has negative consequences for the individual affected, the wider organisation, and patients themselves. The Maslach Burnout Inventory-Human Services Survey (MBI-HSS) has been widely used to assess the prevalence of burnout within oncology services. The MBI-HSS is a self-report measure comprised of three sub-scales - Emotional Exhaustion (EE), Depersonalisation (DP), and Personal Accomplishment (PA). This paper reports on the first study to investigate the psychometric properties of the MBI-HSS when administered to paediatric oncology staff. Two hundred and three paediatric oncology staff recruited through seven UK Principal Treatment Centres (PTCs) and a children’s cancer charity completed the MBI-HSS. The factor structure of the instrument was tested using confirmatory and exploratory factor analysis, with Rasch analysis applied to assess whether the measure meets the requirements of an interval level scale. Cronbach alpha was used to assess internal reliability. Factor analysis did not support the traditional three factor structure of the MBI-HSS but instead suggested seven factors. Rasch analysis and alpha coefficients indicated that while the EE and the PA subscales fulfilled the requirements of an interval level measure for group level diagnosis, DP did not. Further investigation revealed a ‘floor effect’ on many DP items. Whereas the EE and PA subscales of the MBI-HSS can be used in research with paediatric oncology staff working in PTCs, there are considerable problems with the DP subscale and researchers should be cautious in interpreting data from this subscale.

**Introduction**

‘Burnout’ first gained prominence in the 1970s when Herbert Freudenberger wrote about the loss of motivation and commitment to work that he and his colleagues experienced due to encountering excessive emotional and interpersonal workplace stressors (Freudenberger, 1974). Today there is concern about the level of burnout within health services since it is known to be associated with intention to leave the job/specialism, sickness absence, and presenteeism, as well as with patient dissatisfaction, medical error, and poor quality care (for example, Argentero, Dell’Olivo, & Ferretti, 2008; Borritz et al., 2010; Demerouti, Le Blanc, Bakker, Schaufeli, & Hox, 2009; Estryn-Béhar et al., 2007; Poghosyan, Clarke, Finlayson, & Aiken, 2010; Shanafelt, 2010; Whitford, Nadel, & Fish, 2018). Systematic reviews of the literature have found that burnout is elevated in staff working in oncology services. Furthermore, it has been suggested that the proportion of staff experiencing the problem will continue to rise due to increasing workloads and understaffing. As a result, it is recommended that researchers continue to monitor the situation, taking care to use measures with good psychometric properties so that the data is reliable and can be easily synthesised (Medisauskaite & Kamu, 2017; Trufelli et al., 2008). The importance of using psychometrically sound measures of burnout has also been highlighted by those evaluating staff support interventions. A recent review of this evidence concluded that the lack of high quality intervention studies, including studies which have used validated burnout measures, meant it was impossible to make recommendations on how best to prevent staff burnout (Whitford, et al., 2018).

The most commonly used self-evaluation instrument of burnout is the Maslach Burnout Inventory-Human Services Survey (MBI-HSS) (Kaschka, Korczak, & Broich, 2011; Maslach & Jackson, 1996). Within oncology the MBI-HSS has been used extensively with staff working in adult services (Medisauskaite & Kamu, 2017; Trufelli et al., 2008). There is less research on burnout within paediatric oncology, but here too the MBI-HSS has been the measure of choice (Demirci et al., 2010; Gallagher & Gormley, 2009; Laikopoulou et al., 2008; Roth et al., 2011).

The MBI-HSS is a 22 item instrument comprised of three subscales: - emotional exhaustion (EE: 9 items), depersonalisation (DP: 5 five items) and lack of personal accomplishment (PA: 8 items). Each item is scored by the respondent on a ‘scale’ of 0 to 6, with 0 indicating that the feeling ‘never’ happens and 6 indicating that it happens ‘every day’. The EE subscale assesses feelings of being emotionally overextended and exhausted by one’s work, characterised by mental, emotional and physical exhaustion. The DP subscale measures unfeeling and impersonal responses towards service/care recipients, and is seen as a dysfunctional form of detached concern. Finally, the PA subscale assesses feelings of work-related competence and achievement (Maslach & Jackson, 1996). Scores for each subscale are considered separately: they are not combined into a total score. This is because they are measuring three different dimensions which, while inter-related, are not so highly correlated as to constitute a single construct (Leiter & Maslach, 2016). While some researchers have relied on the EE sub-scale as a proxy measure of burnout, considering it to be the primary element of burnout, the scale authors have always argued that burnout is more than just chronic fatigue. To be experiencing burnout people must also be experiencing high levels of depersonalisation and low levels of personal accomplishment (Leiter & Maslach, 2016; Maslach, Leiter, and Schaufeli, 2008).

In relation to scoring the MBI-HSS, the authors advocate that for research purposes the subscale scores are treated as continuous variables. While numerical cut-off points identifying people with high, moderate and low levels of burnout have been provided in the past, these cut-off points have been removed from the most recent edition of the MBI manual (Maslach, Jackson, and Leiter, 2016). Instead the authors recommend that those wanting to use the MBI for individual assessment purposes should use standardised (z) values to ascertain whether the person is scoring highly on EE, DP, and PA. Depending on the pattern of results across the subscales, individuals are then categorised as having a profile of Engaged, Ineffective, Overextended, Disengaged, or Burnout. An intervention can then be devised in accordance with this profile.

When the MBI-HSS was constructed, psychometric measures were developed on the basis of classical test theory. This included testing whether a measure was uni- or multi-dimensional using exploratory factor or confirmatory factor analysis, and assessing internal consistency using Cronbach’s alpha (Hagquist, Bruce, & Gustavsson, 2009). More recently new approaches have emerged and Item Response Theory (IRT), particularly Rasch analysis, is increasingly used to assess both new measures and existing ones (Hobart, Stfen, Zajicek, & Thompson, 2007; Tennant, & Connaghan, 2007). Rasch analysis tests whether the items in a scale satisfy the requirements of fundamental measurement, such that they can generate interval level data (Van Newby, Conner, & Bunderson, 2009). When data satisfy these expectations, a transformation from the ordinal raw score to an interval level latent estimate is considered legitimate. In addition, Rasch analysis involves testing a number of requirements (for example, testing for monotonicity through disorderd thresholds[[1]](#footnote-1), and group invariance through differential item functioning[[2]](#footnote-2) ). As a result, it is argued that Rasch analysis goes beyond what can be achieved through classical test theory in terms of understanding the psychometric properties of a measure (Petrillo, Cano, Mcleod, & Coon, 2015).

Using traditional methods of assessment, MBI-HSS generally performs well. A review and meta-analysis concluded that overall the data supports the three factor model (Stalker, Harvey, Frensch, Mandell, & Adams, 2008). However, the review authors also noted population-specific differences in the reported factor structure and warned against assuming that the three factor structure holds for all staff groups. Early data on the internal consistency of the MBI-HSS suggested it was adequate for research purposes (α 0.90 for EE; 0.79 for DP; and 0.71 for PA (Maslach and Jackson, 1996) . However, a meta-analysis of coefficient alphas highlighted that the alpha estimates for the PA and DP subscales are well below recommended levels for ‘high stake’ decision-making, with the DP subscale having the greatest variability across studies (Wheeler, Vassar, Worley, & Barnes, 2011).

Despite the fact that the measure has been extensively used to research burnout among oncology staff, we have been unable to locate any studies which have examined the factor structure of the MBI-HSS in this population. Furthermore, we could only locate two studies where the reliability of the subscales are reported. Within adult oncology, one study (Kash, Holland, Breitbart, Berenson, Dougherty, Ouellette-Kobasa, S., et al., 2000) reports reliability similar to those originally reported by the measure developers (Cronbach alpha (α) = 0.90 for EE; 0.76 DP; and 0.73 PA). However, a study of paediatric oncology staff reported much lower figures (α =0.80 for EE; = 0.60 DP; and 0.67 PA) (Liakopoulou et al., 2008). The reliability of the DP and PA sub-scales would concern psychometricians who advocate that a Cronbach’s alpha of 0.70-0.80 is required for research tools comparing groups (Nunnally & Bernstein, 1994), with α 0.95 being desirable for high stakes testing (Bland & Altman, 1997).

This paper reports a psychometric assessment of the MBI-HSS when administered to paediatric oncology staff, presenting data on the MBI-HSS factor structure, internal reliability, and whether it meets Rasch model requirements.

**Methods**

***The dataset***

The data was collected during a project which developed measures to assess the work-related stressors and rewards experienced by paediatric oncology staff, with the MBI-HSS used as one of the comparator measures by which to test the construct validities of the new measures (Mukherjee, Beresford, & Tennant, 2014). Although a number of MBI-HSS items use the term ‘recipients’ to refer to service users, it is not uncommon for this term to be replaced by one more appopriate to the setting in which the measure is be used (Maslach & Jackson, 1981). In this study the term ‘recipient’ was replaced with ‘patient’.

Prior to study commencement, NHS Research Ethics Committee approval was sought and granted (REC ref: 08/H0706/131). Before completing the research questionnaire, which included the MBI-HSS and other psychometric measures, participants were given an information lealfet explaining the research, making it clear that participation was entirely volunatry. Completion of the research questionnaire was taken as an indication of informed consent.

A total of 203 staff were recruited from seven Principal Treatment Centres (PTC) in the UK and from a UK children’s cancer charity (CLIC Sargent) which funds social workers, play therapists and youth worker posts in PTCs. The sample included: nurses (n=115, 56.9%); doctors (n=40, 19.8%); social workers (n=29, 14.3%); and play specialists/youth workers (n=18, 8.7%) (NB Data on the professional background of one non-clincal member of staff was missing from their survey response). Staff worked across a number of settings including: inpatients (n=177, 87.2%); outpatients (n=120, 59.1 %); day units (n=114, 56.2%); Bone Marrow Transplant Units (n=99, 48.8%); and in the community (n=40, 19.7%). As would be expected given the majority of repondents were nurses, more women (n=175, 86.6%) than men were recruited to the study. Respondents had worked in paediatric oncology between 6 months and 36 years (median 7 years, 10 months). This sample size is acceptable for the analyses carried out (Linacre 1994). Further methodological details on the research project are reported elsewhere (Mukherjee, Beresford, & Tennant, 2014).

***Data analysis***

*Confirmatory Factor Analysis.* A confirmatory factor analysis (CFA) using MPlus6 was undertaken (Muthén & Muthén, 1998-2011) to test the scales three factor structure: EE, DP and PA. Should the CFA fail to support the three factor structure, an Exploratory Factor Analysis (EFA) was planned.

*The Rasch model and Rasch analysis.* A core element of this investigation was to test the psychometric properties of MBI-HSS using the Rasch model and analytic approaches arising from that model (For overview see Pallent & Tennant, 2007) (see also: http://www.rasch-analysis.com/rasch-analysis.htm). The RUMM2030 programme (Copyright ©1998-2011, Rumm Laboratory Pty Ltd.) was used. The MBI-HSS manual recommends that the subscales are treated as yielding distinct scores. Rasch analysis was therefore performed on each subscale. Steps in the analysis are outlined below.

1. Initial assessment of whether the data fit the Rasch model using summary fit statistics, namely:
   * Chi-square item-trait statistic[[3]](#footnote-3)
   * Person Separation (reliability) Index (PSI)
2. Further examination of the data, including:
   * testing for disordered thresholds;
   * the individual-item fit3;
   * the individual-person fit3;
   * multi-dimensionality within each proposed domain;
   * local dependency[[4]](#footnote-4);
   * differential item functioning (DIF) according to staff group (nurses, doctors, and non-clinical staff), gender, and years of experience in the service;
   * reliability (Cronbach’s alpha).
3. Where subscales did not meet the requirements of the Rasch model, work was undertaken to improve the fit to the model by:
   * rescoring response options to remove disordered thresholds;
   * creating super-items to reduce local dependency;
   * removing misfitting items or persons as appropriate.
4. Finally, where fit to the Rasch model was achieved, a raw score to interval level transformation was carried out using the Logit Person Ability estimate.

**Results**

***Confirmatory factor analysis***

Confirmatory factor analysis failed to support the three factor structure (Chi Square 532.4; df 206; *p* <0.001; CFI 0.856; TLI 0.838; RMSEA 0.092). Several pairs of items within each subscale were locally dependent. After allowing correlated errors for these, the CFA still failed to support the three factor structure (Chi Square 457.6; df 200; *p* <0.001; CFI 0.886; TLI 0.869; RMSEA 0.083). An exploratory factor analysis (EFA) was therefore conducted.

***Exploratory factor analysis***

Exploratory factor analysis identified a solution with seven factors (Chi Square 90.2; df 83; *p* <0.275; RMSEA 0.021). There was only partial support for three dimensions of burnout as put forward by the Maslach model (Maslach and Jackson, 1996). For the EE subscale, two items loaded on a second factor. However, upon inspection, this appeared to be a local dependency issue, as the two items ‘working with people all day is really a strain for me’ and ‘working with people directly puts too much stress on me’ are very similar. Given this, it is possible that the EE subscale is valid. For the DP subscale items, the item ‘I feel patients blame me for some of their problems’ did not load on this subscale, but onto the second EE factor. The DP scale was therefore not supported by this analysis.

***The Rasch analysis***

Prior to undertaking Rasch analysis, a log-likelihood ratio test was performed. Results indicated that the Partial Credit Model (Masters, 1982) was more appropriate than the Rating Scale Model (Andrich, 1978) for this dataset and this was used throughout. Output from analyses and manipulation of the dataset, in terms of fit to the Rasch model, is set out in Table 1 and described below.

*The Emotional Exhaustion (EE) subscale (Table 1, Analyses 1 and 2).*Initial indications were that the EE-subscale items were close to fitting the Rasch model with a chi-square value of 35.51, *p*=0.008. There were a few disordered thresholds but no item misfit. However, there was evidence of multi-dimensionality, with the lower confidence interval for the test of unidimensionality greater than .05. Further investigation revealed that this might be due to local dependency between some of the items. In order to deal with this, testlets combining locally dependent items were created and the model fit was retested (Analysis 2). This indicated fit to the Rasch model (Chi-square interaction 22.94, *p*=0.028). Given this a raw ordinal score to interval transformation was carried out (see Table 2). In the final model, the Cronbach’s alpha was 0.80, and the Person Separation Reliability Index 0.81.

Uniform Differential Item Functioning (DIF) by gender was present in item 14 (‘I feel I am working too hard in my job’ (F=13.75, *p* =.000), with males consistently tending to score higher on this item when their overall EE score is held constant. There was also uniform DIF by staff group on this item (F=16.49,*p* =.000). However, given that in this sample a large proportion of doctors were male, and nurses and non-clinical staff predominantly female, this might be explained by gender. Non-uniform DIF by staff group was also found in one of the sub-tests (Items 2 and 3 combined in a testlet).

*Depersonalisation subscale (Table 1, Analyses 3 - 5).* Initial examination of the data revealed that the data did not fit the Rasch model (Table 1, Analysis 3), with a Chi-square interaction of 38.35, *p*=.000 . There was also low reliability (Cronbach’s alpha = 0.65) and poor ability to distinguish between type of respondent (PSI = 0.32). Examination of individual items revealed that almost all had disordered thresholds. Despite a number of attempts at collapsing the response categories (Analysis 4: data rescored to 0,1,1,2,2,3,3; Analysis 5: data rescored to 0,1,1,1,2,2,2 ), thresholds remained disordered and fit to the model was poor. Given that all but one of the items (Item 15) were misfitting, efforts were made to reduce misfit to the Rasch model by removing Item 15 to see if the remaining items would form a subscale. This too failed to improve fit to the model and the reliability of the data remained very low (Analysis 6). Furthermore, examination of the data revealed a ‘floor effect’ for all of the items on this subscale (Table 3), with the majority of respondents endorsing response category 1 (never) and failing to endorse category 7 (every day) (Figure 1). This pattern is particularly evident in relation to item 5 (‘I feel I treat some patients as if they were impersonal objects’) and item 15 (‘I don’t really care what happens to some patients’).

*Personal Accomplishment subscale (Table 1, Analyses 7-10).* An initial examination of the summary fit statistics (Table 1, Analysis 7) revealed that although the item and person fit was broadly acceptable, and reliability was good, the data did not fit the Rasch model with a Chi-square interaction= 51.40, *p* =.000. Furthermore, there were disordered thresholds on all but two items (Items 9 and 12). A number of attempts were then made to collapse the response categories (Analysis 8 data rescored: 0,1,2,2,2,3,4; Analysis 9 data rescored: 0,1,1,2,2,2,3), but this failed to improve fit to the model. There was also evidence of local dependency between two items (Item 18 - ‘I feel exhilarated after working closely with my patients’; Item 19 - ‘I have accomplished many worthwhile things in this job’). It was therefore decided that these two items should be combined to form a super-item. Following this adjustment (Analysis 10), and retaining rescoring structure from Analysis 9 (0, 1, 1, 2,2,2,3), the data did fit the Rasch model and was found to be unidimensional. A transformation from the ordinal raw score to interval level equivalent was therefore carried out (see Table 4). There was no evidence of DIF on any items in this subscale. The final model had a Cronbach’s alpha of 0.70 and a PSI of 0.71.

**Discussion**

The MBI - HSS is one of the most widely used measures of burnout, administered across countries and to a wide range of professional groups. However, concerns have been raised about its reliability with specific populations. Within oncology, it has been used within paediatric and adult services despite the fact that its psychometric properties have not been tested in these populations. The work reported in this paper is, we believe, the first attempt to examine the properties of the MBI-HSS with respect to staff working in paediatric oncology, using both traditional and more modern approaches to psychometric testing.

Our analyses found no evidence to support a three factor structure of the MBI-HSS when it is administered to paediatric oncology staff. When an exploratory factor analysis was performed, the simplest structure obtained comprised seven factors. However, these factors had little conceptual coherence, which is unsurprising given that the MBI-HSS is only comprised of 22 items.

Examining the performance of MBI-HSS using Rasch analysis, we found mixed support for the measure’s reliability. On a positive note, after dealing with local dependency, the EE subscale did fit the Rasch model expectations and could be considered a uni-dimensional scale, providing interval level measurement when required. The standard approach to dealing with local dependency is either by *a priori* conceptual groupings (e.g. sub-domains) or post-hoc groupings following the evidence from the residual correlation analysis, and when the grouping makes conceptual sense, which is the case here. Furthermore, the seven category response option worked well for this subscale, with little evidence of disordered thresholds. Results from the DIF analyses indicate it is appropriate for use with clinical and non-clinical staff, males and females, and staff who have worked in the service for different periods of time. However, it should be noted that some DIF (e.g. by gender) was observed, although no action was taken to resolve this. Replication of the analysis elsewhere, giving the same results for DIF would suggest the need to consider how best to resolve this problem. It should also be noted that the value of Cronbach’s alpha (α 0.80) indicates that this subscale should only be used at a population level in order to compare groups, rather than as a diagnostic tool (Bland & Altman, 1997).

Initial examination of the PA subscale found a number of items had disordered thresholds. This issue was resolved by changing the scoring structure so that the central response options were collapsed into one category (i.e. response categories 0,1,2,3,4,5,6 were rescored 0,1,1,2,2,2,3 respectively). Once rescored the data did fit the Rasch model. However, as with the EE subscale, a Cronbach’s alpha of 0.70 indicates this subscale cannot be used for individual assessment.

Finally, initial examination of the DP subscale revealed poor fit to the Rasch model, low reliability and a poor ability to distinguish between groups. Examination of individual items revealed that almost all had disordered thresholds which could not be corrected by collapsing response categories. Removing items from the subscale did not improve fit to the model. On reflection, these findings are perhaps unsurprising given many of the items in this subscale represent relatively extreme examples of depersonalisation (e.g. ‘I don’t really care what happens to some patients’ and ‘I feel I treat some patients as though they are impersonal objects’). Another key issue emerging from our analyses of the DP subscale was that floor effects were observed for many items. This could indicate that either few staff experience these relatively ‘extreme’ indicators of depersonalisation, or there may be a social desirability effect with few staff willing to admit such feelings. These findings align with previous investigations into the performance of the MBI-HS in which researchers report problems with the reliability, the factor structure and/or the face validity of the DP subscale when administered to staff working in some health, social care or welfare settings (Chao, McCallion, & Nickle, 2011; Kristensen et al., 2005; Schaufeli & Taris, 2005; Stalker et al., 2008). Indeed, problems with the DP subscale is one reason why some researchers have developed new measures of job burnout (Kristensen et al., 2005) or recommend using alternatives (Schaufeli & Taris, 2005).

The failure of the DP subscale to meet the requirements of the Rasch model raises a problem for those wishing to use the MBI-HSS to assess burnout in the paediatric oncology workforce. Whilst EE is the most widely reported and analysed aspect of burnout, and seen by some as the core aspect of the phenomenon (e.g. Kristensen et al., 2005, Shirom, 2003), Maslach and colleagues have always argued that burnout is a multi-dimensional syndrome and not simply ‘fatigue’ or ‘chronic exhaustion’ (Leiter & Maslach, 2016; Maslach, Schaufeli, & Leiter 2001; Schaufeli & Taris, 2005; Schaufeli et al. 2009). For many researchers depersonalisation is of particular interest since this aspect of burnout is thought to have most direct consequences for the healthcare provider-patient relationship and the quality of care patients receive (Firth-Cozens & Cornwell, 2009). An alternative explanation is that paediatric oncology staff are a highly motivated workforce not affected by depersonalisation which is resulting in them scoring at the floor of this subscale. If this is the case then the low reliability of the measure may be due to the homogeneity of the group. Clearly, the reliability of the DP subscale needs further investigation.

In summary, the findings reported here confirm the relatively robust psychometric properties of the EE and PA subscales of the MBI-HSS when used with paediatric oncology staff working in UK PTCs. However, the data should only be used for group level comparison. Furthermore, while the EE subscale can be scored using the original 7 category options, the PA subscale should be scored using a 4 category response option (0,1,1,2,2,2,3). The DP subscale, in contrast, failed to meet Rasch model expectations, and a floor effect was found for many of the items. It is questionable, therefore, whether this subscale of the MBI-HSS should be used with (at least) UK paediatric oncology staff working in PTCs. Where people do choose to use this subscale we recommend caution in the interpretation of the results. Furthermore, it is important to note that the transformation tables are only valid for complete data. Replications of this work – including across different countries - are required.

**Acknowledgements**

This paper is based on data collected as part of the Life in Paediatric Oncology Project (LiPOP), a research project aimed at developing and validating a measure of the work-related stressors and rewards experienced by paediatric oncology staff working in principal treatment centres. We are grateful to all the staff who took the time to participate in LiPOP.

**Funding**

This paper used data collected as part of LiPOP, a research project funded by Cancer Research UK [grant number C28901/A10114]

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Supplementary materials**

Any queries or data requests should be submitted to the corresponding author for consideration. Access to available anonymised data may be granted following review.

**References**

Andrich, D. (1978) Rating Formulation for ordered response categories. Psychometrica, 43,

561-573.

Argentero, P., Dell’Olivo, B., & Ferretti, M. (2008) Staff Burnout and Patient Satisfaction with the Quality of Dialysis Care. American Journal of KidneyDiseases*,* 51(1), 80-92.

Bland, J. & Altman, D. (1997) Statistics notes: Cronbach’s alpha, BMJ, 314, 572.

Borritz, M., Christensen, K., Bültmann, U., Rugulies, R., Lund, T., Andersen, I., et al. (2010)

Impact of Burnout and Psychosocial Work Characteristics on Future Long-Term Sickness Absence. Prospective Results of the Danish PUMA Study Among Human Service Workers. Journal of Occupational and Environmental Medicine, 52 (10), 964-70. [https://doi: 10.1097/JOM.0b013e3181f12f95.](https://doi:%2010.1097/JOM.0b013e3181f12f95.)

Chao,S., McCallion, P., & Nickle, T. (2011) Factorial validity and consistency of the Maslach

Burnout Inventory among staff working with persons with intellectual disability and dementia. Journal of Intellectual Disability Research, 55(5), 529-36. https://doi: [10.1111/j.1365-2788.2011.01413.x](https://doi.org/10.1111/j.1365-2788.2011.01413.x)

Christensen, K.B., Makransky, G. & Horton, M. (2017) Critical values for Yen’s Q 3:

identification of local dependence in the Rasch model using residual correlations. Applied Psychological Measurement, 41(3), 178-194. https://[doi.org/10.1177/0146621616677520](https://doi.org/10.1177%2F0146621616677520)

Demerouti, E., Le Blanc, P., Bakker, A., Schaufeli, W., & Hox, J. (2009) Present but sick: a

three‐wave study on job demands, presenteeism and burnout. Career Development International, 14 (1), 50-68. [https://doi: 10.1108/13620430910933574](https://doi:%2010.1108/13620430910933574)

Demirci, S., Yildirim, Y., Ozsaran, Z., Uslu, R., Yalman, D., & Aras A. (2010). Evaluation of

burnout syndrome in oncology employees. Medical Oncology*,* 27(3), 968-74.  https://doi: [10.1007/s12032-009-9318-5](https://doi.org/10.1007/s12032-009-9318-5)

Estryn-Béhar,M., Heijden B., Ogińska, H., Camerino, D., Nézet, O., Conway, P., et al (2007)

The Impact of Social Work Environment, Teamwork Characteristics, Burnout, and Personal Factors upon Intent to Leave among European Nurses. Medical Care, 45(10), 939-50. https:// doi: [10.1097/MLR.0b013e31806728d8](https://doi.org/10.1097/MLR.0b013e31806728d8)

Firth-Cozens, J., & Cornwell, C. (2009). The Point of Care: Enabling Compassionate Care in

Acute Hospital Settings. The King’s Fund, London.

Freudenberger, H. (1974) Staff Burn-Out. Journal of Social Issues, 30 (1):159-65. <https://doi.org/10.1111/j.1540-4560.1974.tb00706.x>

Gallagher, R., & Gormley, D. (2009). Perceptions of stress, burnout, and support systems in

pediatric bone marrow transplantation nursing. Clinical Journal of Oncology Nursing*,*13(6), 681-5.  https://doi: [10.1188/09.CJON.681-685](https://doi.org/10.1188/09.CJON.681-685)

Hagquist, C., Bruce, M., & Gustavsson, J. (2009). Using the Rasch model in nursing research:

an introduction and illustrative example. International Journal of Nursing Studies*,* 46(3), 380-93. https://doi: [10.1016/j.ijnurstu.2008.10.007](https://doi.org/10.1016/j.ijnurstu.2008.10.007)

Hobart, J., Stfen, J., Zajicek, J., & Thompson, A. (2007). Rating scales as outcome measures

for clinical trials in neurology: problems, solutions, and recommendations. Lancet Neurology*,* 6, 1094-105*.* https://doi: [10.1016/S1474-4422(07)70290-9](https://doi.org/10.1016/S1474-4422(07)70290-9)

Kaschka, W., Korczak, D., & Broich, K. (2011) Burnout: a fashionable diagnosis. Deutsches

Arzteblatt, 108 (46), 781-7. [https://doi: 10.3238/arztebl.2011.0781.](https://doi:%2010.3238/arztebl.2011.0781.)

Kash, K., Holland, J., Breitbart, W., Berenson, S., Dougherty, J. Ouellette-Kobasa, S., et al.

(2000). Stress and burnout in oncology. Oncology, 14 (11), 1621-33.

Kristensen,T., Borritz, M., Villadsen, E., & Christensen, K. (2005) The Copenhagen Burnout

Inventory: A new tool for the assessment of burnout. Work & Stress,19 (3), 192-207. <https://doi.org/10.1080/02678370500297720>

Laikopoulou, M., Panaretaki, I., Papadakis, V., Katsika, A., Sarafidiou, J., Laskari, H., et al.

(2008). Burnout, staff support and coping in pediatric oncology. Supportive Care and Cancer, 16, 143-50. https://doi:[10.1007/s00520-007-0297-9](https://doi.org/10.1007/s00520-007-0297-9)

Leiter. M., & Maslach, C. (2016) Latent burnout profiles: A new approach to understanding

the burnout experience. Burnout Research, 3, 89-100. https://[doi.org/10.1016/j.burn.2016.09.001](https://doi.org/10.1016/j.burn.2016.09.001)

Linacre, J. (1994) Sample size and item calibration stability. Rasch Meas Transac, 7, 328.

Maslach, C. & Jackson, S. (1981) The measurement of experienced burnout. Journal of

Organizational Behavior, 2 (2), 99-113. <https://doi.org/10.1002/job.4030020205>

Maslach, C. & Jackson,S. (1996) *Maslach Burnout Inventory Manual*, Second Edition,

Consulting Psychologists Press, Inc., USA.

Maslach, C., Jackson,S. and Leiter, M. (2016) *Maslach Burnout Inventory Manual,* Fourth

Edition, Mind Garden, Inc.

Maslach,C., Schaufeli, W., & Leiter, M. (2001) Job burnout. Annual review of psychology,

52:397-422. [https://doi.org/10.1146/annurev.psych.52.1.397](https://doi.org/10.1146/annurev.psych.52.1.397%20)

Masters,G. (1982) Rasch Model for partial credit scoring. Psychomertica, 47, 149-174.

Mukherjee, S., Beresford, B., & Tennant, A. (2014) Staff burnout in paediatric oncology: new tools to facilitate the development and evaluation of effective interventions. European Journal of Cancer Care, 23 (4), 1365-2354. <https://doi.org/10.1111/ecc.12176>

Muthén, L. K., & Muthén, B. O. (1998-2011). Mplus User's Guide. Sixth Edition. Los Angeles,

CA: Muthén & Muthén.

Nunnally J.C., & Bernstein I.H. (1994) Psychometric Theory (3rd ed.). New York: McGraw-

Hill.

Petrillo, J., Cano, S., Mcleod, L., & Coon, C. (2015) Using classical teat theory, item response

theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. Value in Health, 18, 25-34.

Poghosyan, L., Clarke, S., Finlayson, M., & Aiken, L. (2010) Nurse Burnout and Quality of

Care: Cross-National Investigation in Six Countries. Research in Nursing & Health, 33(4), 288-298. [https://doi: 10.1002/nur.20383.](https://doi:%2010.1002/nur.20383.)

Roth, M., Morrone, K., Moody, K., Kim, M., Wang, D., Moadel, A., et al. (2011). Career

burnout among pediatric oncologists. Pediatric Blood and Cancer, 57,1168-73[.  https://doi: 10.1002/pbc.23121.](.%20 https:/doi:%2010.1002/pbc.23121.)

Schaufeli WB, Leiter MP, & Maslach C. (2009) Burnout: 35 years of research and practice.

Career Development International, 14(3), 204-20. https://[doi.org/10.1108/13620430910966406](https://doi.org/10.1108/13620430910966406)

Schaufeli, W.B. & Taris, T.W. (2005) The conceptualization and measurement of burnout:

Common ground and worlds apart. Work & Stress, 19, 356-362. [https://doi: 10.1080/02678370500385913](https://doi:%2010.1080/02678370500385913)

Shanafelt, T. (2010) Burnout and medical errors among American surgeons. Annals of

surgery, 251(6), 995-1000[. https://doi: 10.1097/SLA.0b013e3181bfdab3.](.%20https:/doi:%2010.1097/SLA.0b013e3181bfdab3.)

Shirom, A. (2003). Job-related burnout: A review. In J.C. Quick & L.E. Tetrick, Handbook of

Occupational Health Psychology. Washington, DC: American Psychological Association, 245-264. https://doi:  [10.1037/10474-012](http://dx.doi.org/10.1037/10474-012)

Stalker, C., Harvey, C., Frensch, K., Mandell, D., & Adams, G. (2008) Confirmatory Factor

Analysis of the Maslach Burnout Inventory. Journal of Public Child Welfare, 1(3),77-94. [https://doi: 10.1300/J479v01n03\_05.](https://doi:%2010.1300/J479v01n03_05.)

Tennant, A., & Conaghan, P. (2007) The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied and what should one look for in a Rasch paper? Arthritis Rheumatology, 57,1358-62. https://doi: [10.1002/art.23108](https://doi.org/10.1002/art.23108)

Pallant, J., & Tennant, A. (2007) An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). British Journal of Clinical Psychology, 46 (1), 1-18. <https://doi:10.1348/014466506X96931>

Trufelli D.C., Bensi C.G., Garcia J.B., Narahara J.L., Abrao M.N., Diniz R.W., et al. (2008)

Burnout in cancer professionals: a systematic review and meta-analysis. Eur J CancerCare, 17(6):524-31. [https://doi: 10.1111/j.1365-2354.2008.00927](https://doi:%2010.1111/j.1365-2354.2008.00927)

Van Newby A., Conner G.R., & Bunderson C.V. (2009). The Rasch model and additive

conjoint measurement. Journal of Applied Measurement, 10, 348-354. https://doi: [10.1016/j.jval.2014.10.005](https://doi.org/10.1016/j.jval.2014.10.005)

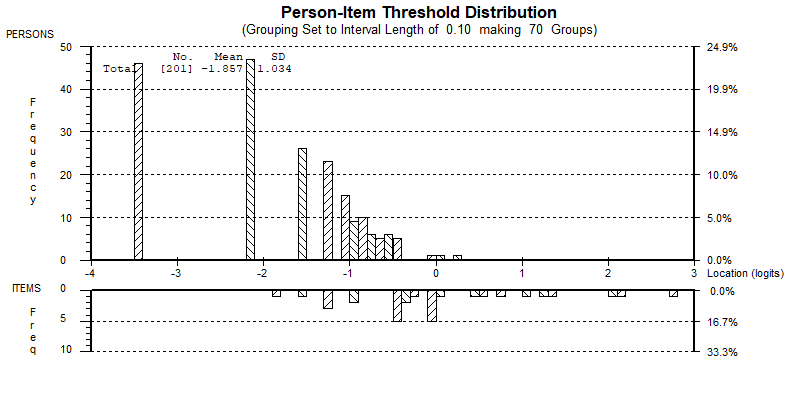
Wheeler, D., Vassar, M., Worley, J., & Barnes, L. A. (2011) Reliability Generalization Meta-

Analysis of Coefficient Alpha for the Maslach Burnout Inventory. Educational and Psychological Measurement, 71 (1), 231-44. <https://doi.org/10.1177%2F0013164410391579>

Whitford, B., Nadel, A.L, & Fish, J.D. (2018) Burnout in pediatric hematology/oncology- time

to address the elephant by name. Paediatric Blood and Cancer, 65:e27244. [https://doi: 10.1002/pbc.27244](https://doi:%2010.1002/pbc.27244)

**Figure 1.** Distribution of the Depersonalisation Subscale



|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1.** Summary of Rasch Analysis | | | | | | | | |
| Analysis | Sub-scale | Item Residual | Person Residual | χ2 Interaction | | PSI | Cronbach’s alpha | Unidimensionality |
|  |  | Mean  (SD) | Mean (SD) | Value | *p* |  |  | Percentage sig at <5% (95% CI) |
| 1 | MBI-HSS EE | -0.71  (1.20) | -0.42 (1.01) | 35.51 | 0.008 | 0.845 | 0.84 | 8.96 (2.87-15.05) |
| 2 | MBI-HSS EE | -0.58  (1.41) | -0.43 (1.06) | 22.94 | 0.028 | 0.810 | 0.80 | 6.47 (0.38 - 12.56) |
| 3 | MBI-HSS DP | -1.26  (0.80) | -0.37 (0.49) | 38.35` | 0.000 | 0.324 | 0.65 | 0 (-6.09 - 6.09) |
| 4 | MBI-HSS DP | -0. 68  (1.31) | -0.33 (0.63) | 37.95 | 0.000 | 0.321 | 0.64 | 0 (-6.09 - 6.09) |
| 5 | MBI-HSS DP | -0.27  (1.64) | -0.31 (0.70) | 35.90 | 0.000 | 0.349 | 0.63 | 0.5 (-5.59 - 6.59) |
| 6 | MBI-HSS DP | 0.06  (1.73) | -0.27 (0.82) | 34.598 | 0.000 | 0.224 | 0.60 | Could not be calculated |
| 7 | MBI-HSS PA | 0.05  (1.63) | -0.32 (1.04) | 51.403 | 0.000 | 0.764 | 0.74 | 3.98 (-2.11-10.07) |
| 8 | MBI-HSS PA | 0.29  (1.33) | -0.33 (1.20) | 43.489 | 0.000 | 0.755 | 0.75 | 4.98 (-1.12-11.07) |
| 9 | MBI-HSS PA | -0.06  (1.45) | -0.37 (1.22) | 36.674 | 0.002 | 0.733 | 0.75 | 3.00 (-3.09-9.09) |
| 10 | MBI-HSS PA | -0.23  (1.61) | -0.39 (1.16) | 24.901 | 0.035 | 0.707 | 0.70 | 3.00 (-3.09-9.09) |
| Requirements for satisfactory fit | | 0.0  (<1.4\*) | 0.0 (<1.40) |  | >0.01 | ≥0.7 | ≥0.7 | Lower CI (<.0.5) |
| SD: Standard Deviation; χ2: Chi-square; PSI: Person Separation Index; CI: Confidence Interval; \* May be inflated when testlets are used. | | | | | | | | |

**Table 2.** Raw Score Interval Scale Transformation for Emotional Exhaustion Subscale

|  |  |  |  |
| --- | --- | --- | --- |
| Raw (ordinal) score | Transformation to  interval Score | Raw (ordinal) score | Transformation to interval score |
| 0 | 0.0 | 28 | 31.5 |
| 1 | 6.3 | 29 | 31.9 |
| 2 | 10.4 | 30 | 32.2 |
| 3 | 13.1 | 31 | 32.5 |
| 4 | 15.2 | 32 | 32.8 |
| 5 | 16.8 | 33 | 33.2 |
| 6 | 18.1 | 34 | 33.5 |
| 7 | 19.3 | 35 | 33.9 |
| 8 | 20.3 | 36 | 34.3 |
| 9 | 21.2 | 37 | 34.6 |
| 10 | 22.1 | 38 | 35.1 |
| 11 | 22.9 | 39 | 35.5 |
| 12 | 23.6 | 40 | 36.1 |
| 13 | 24.3 | 41 | 36.6 |
| 14 | 24.9 | 42 | 37.2 |
| 15 | 25.6 | 43 | 37.9 |
| 16 | 26.1 | 44 | 38.7 |
| 17 | 26.7 | 45 | 39.5 |
| 18 | 27.2 | 46 | 40.4 |
| 19 | 27.8 | 47 | 41.3 |
| 20 | 28.2 | 48 | 42.2 |
| 21 | 28.7 | 49 | 43.2 |
| 22 | 29.2 | 50 | 44.3 |
| 23 | 29.6 | 51 | 45.5 |
| 24 | 30.0 | 52 | 47.2 |
| 25 | 30.4 | 53 | 49.8 |
| 26 | 30.8 | 54 | 54.0 |
| 27 | 31.2 |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 3.** Category Response Frequencies for Depersonalisation Subscale. | | | | | | | | | |
| Item | | Response category (N=203) | | | | | | | |
|  |  | Never | A few times a year | Once a month or less | A few times a month | Once a week | A few times a week | Every day | Missing |
| 5 | I feel I treat some patients as if they were impersonal objects | 163 | 25 | 4 | 1 | 4 | 3 | 0 | 3 |
| 10 | I have become more callous toward people since I took this job | 135 | 41 | 10 | 9 | 2 | 4 | 0 | 2 |
| 11 | I worry that this job is hardening me emotionally | 97 | 63 | 15 | 10 | 3 | 5 | 7 | 3 |
| 15 | I don’t really care what happens to some patients | 188 | 10 | 3 | 0 | 0 | 0 | 0 | 2 |
| 22 | I feel patients blame me for some of their problems | 96 | 66 | 16 | 11 | 8 | 2 | 1 | 3 |

**Table 4**. Raw Score Interval Scale Transformation for Personal Accomplishment Subscale (where initial response category scoring of 0,1,2,3,4,5,6 is rescored to 0,1,1,2,2,2,3)

|  |  |
| --- | --- |
| Raw (ordinal) score | Transformation to the interval score |
| 0 | 0 |
| 1 | 2.6 |
| 2 | 4.4 |
| 3 | 5.7 |
| 4 | 6.7 |
| 5 | 7.6 |
| 6 | 8.3 |
| 7 | 9.0 |
| 8 | 9.7 |
| 9 | 10.3 |
| 10 | 10.9 |
| 11 | 11.6 |
| 12 | 12.1 |
| 13 | 12.8 |
| 14 | 13.4 |
| 15 | 14.1 |
| 16 | 14.8 |
| 17 | 15.5 |
| 18 | 16.2 |
| 19 | 17.0 |
| 20 | 17.8 |
| 21 | 18.8 |
| 22 | 20.0 |
| 23 | 21.7 |
| 24 | 24.0 |

1. The test for disordered thresholds is an empirical test of the hypothesis that response categories are in the correct order, that is they represent increasing levels of a latent attribute or trait (monotonicity) . [↑](#footnote-ref-1)
2. Differential Item Functioning (DIF**) *-*** occurs when people from different groups (commonly gender or ethnicity) with the same estimated latent trait have a different probability of giving a certain response on a scale item. [↑](#footnote-ref-2)
3. Misfit of items and persons indicated at p-value of <0.01, or fit residuals outside of the +/- 2.5 range. [↑](#footnote-ref-3)
4. Local dependency is determined by residual correlations of 0.2 above the average residual correlation (see Christensen, K.B., Makransky, G. and Horton, M., 2017) . It implies that items are very similar and do not contribute a full independent items’ worth of information. In some cases, it can imply redundancy in the item set. [↑](#footnote-ref-4)