



Available online at www.sciencedirect.com

ScienceDirect

Health Professions Education 5 (2019) 259–274



www.elsevier.com/locate/hpe

An Examination of the Structural Validity of the Maslach Burnout Inventory-Student Survey (MBI-SS) Using the Rasch Measurement Model

José Manuel Pérez-Mármol^{a,*}, Ted Brown^b

^aDepartment of Physiotherapy (Occupational Therapy), Faculty of Health Sciences, University of Granada, Avenue de la Ilustración, 60, 18016 Granada, Spain

^bDepartment of Occupational Therapy, Faculty of Medicine, Nursing and Health Sciences, Monash University - Peninsula Campus, Frankston, Victoria, Australia

Received 4 February 2018; received in revised form 12 May 2018; accepted 14 May 2018

Available online 19 May 2018

Abstract

Background: Health professional students often experience stress and potential burnout when completing their entry-to-practice education. Therefore, having valid scales such as the Maslach Burnout Inventory-Student Survey (MBI-SS) to detect and monitor burnout amongst health professional students is paramount.

Purpose: This study investigated the structural validity of the three individual Maslach Burnout Inventory-Student Survey (MBI-SS) subscales. In particular, the dimensionality, scalability, differential item functioning (DIF) based on gender, rating scale structure, and reliability of the three MBI-SS were examined.

Method: A sample of 225 Australian occupational therapy undergraduate students completed it. The Rasch Measurement Model (RMM) was used to analyse the MBI-SS item data from the three MBI-SS subscales: Exhaustion, Cynicism, and Professional / Academic Efficacy.

Results: Results supported the unidimensionality and scalability of the three individual MBI-SS and all items from the three subscales met the RMM fit statistics requirements. None of the subscale items exhibited DIF based on gender and the appropriateness of the MBI-SS rating scales were confirmed. The item and person reliability indices were > 0.80 and > 0.77 respectively for each subscale.

Discussion and conclusion: The RMM analysis findings indicate that the three individual MBI-SS subscales exhibited adequate levels of structural validity in relation to dimensionality and DIF when completed by occupational therapy students.

© 2018 King Saud bin AbdulAziz University for Health Sciences. Production and Hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Burnout; Validity; Education; Educational measurement; Occupational therapy

*Corresponding author.

E-mail address: josemapm@ugr.es (J.M. Pérez-Mármol).

Peer review under responsibility of AMEEMR: the Association for Medical Education in the Eastern Mediterranean Region

1. Introduction

Burnout was first mentioned as an issue occurring in the helping professions. Bradley in 1969, Herbert Freudenberger in 1974, and Cristina Maslach in 1976 are the authors who first coined this term.¹ Maslach and Jackson² defined burnout as a syndrome of *emotional exhaustion*, *cynicism*, and *reduced personal accomplishment* that is frequently manifested among individuals who do some kind of ‘people-work’.^{1,2} Burnout as a theoretical construct has been linked with human services workers and health care professionals, but currently research includes other areas of employment and other groups of participants.^{3,4} Several studies have been completed involving occupational therapists that have identified several factors that may contribute to therapist burnout including excessive demands on time; perceived work role overload; work conflicts; lack of autonomy, respect, resources, supervision and continuing professional development opportunities; inadequate remuneration and career progression structure; role confusion; role ambiguity; low psychological detachment from work during out-of-work hours (i.e. being able to ‘shut off’ from work outside work hours); difficulties being able to say “no” in specific work situations; having more than ten years of experience; or managing multiple dimensions related to patient care.^{5–17}

The interest in examining burnout has also extended to undergraduate students enrolled in professional-oriented and health-related entry-to-practice tertiary education courses that provide assistance to people (i.e. pharmacy, physiotherapy, medicine, dentistry, social work, nursing).^{18–20} Although stress reported by occupational therapy undergraduate student has been investigated previously,^{21,22} to our knowledge, no studies have been published to date in the peer-reviewed literature that involved the self-reported levels of ‘burnout syndrome’ in occupational therapy undergraduate students. In previously published studies, occupational therapy students have reported high levels of stress, expressing negative emotions such as burden, overload, and concerns regarding course expectations, examinations, amount of class work, lack of free time, and long hours of study.^{21–23} Hence, both clinical and academic educators need to be aware of the potential stress and burnout levels of occupational therapy students, so as to be able to respond to these issues in a timely, appropriate, and constructive manner. If stress and burnout amongst occupational therapy students is going to be monitored, then it is important to have valid and reliable scales that measure this construct.

The Maslach Burnout Inventory-Student Survey (MBI-SS) has been a frequently used measure of burnout in university students previously.^{24–29} The dimensions assessed in the MBI-SS correspond with feeling exhausted due to study demands (i.e. Exhaustion), showing a cynical, uninvolved attitude towards one’s own studies (i.e. Cynicism), and feeling incompetent in one’s role as a student (i.e. Professional / Academic Efficacy).²⁸ It is essential that the validity and reliability of scales used to assess levels of burnout in students (such as the MBI-SS) be examined to ensure that the dimensions they claim to measure are done so in a rigorous manner.^{30,31} The validity of the MBI-SS has been examined using Classical Test Theory factor analytic approaches previously, but it has not been investigated using Item Response Theory approaches.^{27,29,32–37} The MBI-SS has also been translated in several different languages including French, Serbian, Spanish, German, Japanese, and Korean which indicates its international usage.^{27,29,34–39}

The Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) proposal specified that evidence of six types of measurement properties should be reported about health-related measures and scales: internal consistency, reliability, measurement error, content validity, construct validity, and criterion validity.⁴⁰ According to the COSMIN, construct validity evidence should report details of a scale’s structural validity, hypothesis testing, and cross-cultural validity.⁴¹ Structural validity is defined as the degree to which scores of a scale are an adequate indication of the dimensionality of the construct or attribute being measured.^{41–43} Evidence of the structural validity of a scale can take the form of differential item functioning studies and dimensionality studies.⁴² One type of statistical model that can be used to examine aspects of the structural validity of a scale is the Rasch Measurement Model (RMM).⁴⁴

The present study investigated the structural validity of the three individual MBI-SS subscales using the RMM.⁴⁴ The RMM is a type of Item Response Theory that generates ordinal level scores.⁴⁵ With the RMM, the unidimensionality of scales used to measure constructs, the interval level scaling (also referred to as scalability), the rating scale structure of instruments, if differential item functioning exists in a scale, and its reliability can be investigated.^{44,46} Unidimensionality refers how well the items of a scale represent a single underlying dimension or construct using a mathematical model to know how well they fit Rasch model expectations.⁴⁷ Interval level scaling linearizes the nominal raw items scores into a logit scale. The key

assumption is that the units of a measure should be equal quantities across the range of the construct under consideration.^{20,47,48} Goodness-of-fit statistics are used to determine how well the data from the items and participants fit the expectations of the RMM. DIF is said to occur when one group of participants, based on gender, age or other similar variables, respond in a different manner to individual scale items measuring the same underlying construct.⁴⁵ These analyses may provide valuable information for prospective users of the MBI-SS who may want to apply it with occupational therapy students to monitor their burnout levels.

Academic and fieldwork educators could use the MBI-SS to monitor the burnout levels of students or the MBI-SS could be used to evaluate an intervention designed to ‘stress-proof’ or minimize the impact of stress on students. Similarly, with access to a valid scale such as the MBI-SS, further research can be completed to investigate the risk factors of burnout syndrome or to assess different approaches for addressing this problem in practice and education settings. For these reasons, the purpose of the study was to investigate the structural validity and reliability of the three individual MBI-SS subscales using the RMM when completed by a sample of Australian occupational therapy undergraduate students. Specifically, the unidimensionality, scalability, differential item functioning, rating scale structure, and reliability of the three individual MBI-SS subscales were investigated.

2. Method

2.1. Research design

A cross-sectional study design was used to investigate the measurement properties of the three subscales of the MBI-SS.

2.2. Participants

The questionnaires were completed by 225 Australian occupational therapy undergraduate students using a convenience sampling approach. The students were enrolled in the first, third and fourth year the Bachelor of Occupational Therapy (Honours) course that is offered at Monash University – Peninsula Campus, Frankston, Victoria, Australia. The study inclusion criteria were being an occupational therapy student enrolled at Monash University in 2015 and consenting to take part in the study. Second year students were not able to be recruited as part of the sample since they

were not accessible to complete the questionnaire due to the scheduling of their tutorials and lectures.

Monash University Human Research Ethics Committee approval was obtained on the 9 April 2015 (project number: 2015-6069-5898). This study met the requirements of the National Statement on Ethical Conduct in Human Research and was conducted in accordance with the Declaration of Helsinki 2013.

2.3. Instrumentation

A demographic questionnaire was used to collect relevant information (such as age, gender, current degree level and weekly time allocation for study) about the participant group. Data about burnout reported by students was obtained using the MBI-SS. The MBI-SS consists of 15 items that are allocated to one of three subscales: *Exhaustion* with five items, *Cynicism* with four items, and *Professional / Academic Efficacy* with six items.²⁸ The MBI-SS uses a seven-point frequency rating scale that ranges from 0 (never) to 6 (every day) for participants to rate each item. High scores on the Exhaustion and Cynicism subscales and low scores on the Professional / Academic Efficacy scale are interpreted as indicating the presence of burnout in respondents.²⁸

2.4. Procedure

Undergraduate students were approached at the conclusion of a lecture during Semester One 2015 by a non-teaching member of staff and were asked to complete the MBI-SS. Consent was inferred if students volunteered to complete and return the questionnaire. It took participants on average 10 min to complete the questionnaire.

3. Data analysis

The RMM analysis procedure was used to investigate each of the three MBI-SS subscales using the Winsteps computer program version 3.73.⁴⁹ The Winsteps computer program runs Rasch analysis that is based on Item Response Theory, in contrast with other statistical analysis software such as SPSS program that executes analyses from a classical approach for validation of scales. The RMM analysis is an iterative process, with the objective of reaching the best fit of the data to the model by testing the model's assumptions.^{44,50} The RMM was used to examine the main features of the MBI-SS such as scalability of the three

subscales, test unidimensionality, differential item functioning, rating scale structure, and reliability.

3.1. Scale unidimensionality

Item goodness-of-fit statistics generated from the Winsteps computer program was used to analyze the dimensionality of the three MBI-SS subscales.⁴⁴ Principal component analysis (PCA) of item residuals identifies the primary construct under consideration (also referred to as the principle component) and it must account for at least 60% of the total variance and the eigenvalue in the first contrast should be less than 2.0.^{46,49}

3.2. Scalability/interval level scaling

The assumption for this property is that the units of measure reflect equal quantities across the range of the construct. Raw items scores obtained from the three MBI-SS subscales are ordinal and they must be linearized so that they correspond to equal increments of the underlying construct. Linearize refers to execute a linear transformation to convert and calibrate Rasch parameter estimations to the same scale. Those transformed estimations can be used to provide score equivalents between the raw and scaled scores of the different test forms.^{51,52} The RMM has two types of goodness-of-fit statistics (infit and outfit) to evaluate how well test items fit with the expectations of the model. These statistics are exhibited as mean square (MnSq) and standardized Z values (Zstd). For MnSq values, the commonly agreed range is from 0.71 to 1.3 and for the Zstd values from -2 to $+2$.⁴⁴

An item is considered to misfit Rasch Model expectations when the MnSq and Zstd values are outside of these score-fit ranges. Infit and outfit statistics into MnSq are calculated in a different manner; infit values confers more relevance to the ability scores of participants closer to the item value, that is, the closer ability respondents to the item's difficulty are, the more sensitive insight into item's performance; outfit statistic is more sensitive to the impact of outlying scores because it is not weighted.⁴⁴

It is also important to take into account the basic rating scale assumptions to analyze the appropriateness of the seven-point rating scales of the three subscales of MBI-SS.⁵³ As Linacre⁵⁴ recommended, Rasch-Andrich thresholds were used to explore the

Table 1

Demographic characteristics of the undergraduate Monash student sample ($N = 225$).

	Frequency	Percentage (%)
Gender		
Female	200	88.90
Male	25	11.10
Year level of enrolment		
1st year	97	43.10
3rd year	71	31.60
4th year	57	25.30
Enrolment status		
Full-time	215	95.60
Part-time	10	4.40
Entrance to education program directly from high school?		
Yes	112	49.80
No	113	50.20
	Mean (SD)	
Average number of hours of direct occupational therapy class time attended per week	15.85 (6.66)	
Average number of hours of indirect time spent working on occupational therapy studies	15.19 (8.03)	
Average number of hours per week students spend working at a paid job while completing occupational therapy studies	10.90 (8.05)	
Average number of hours occupational therapy students spend per week engaged in self-care activities	16.32 (16.31)	

Note: SD = Standard Deviation.

appropriateness across individual items in the seven-level rating scale.⁵¹ The appropriateness across individual items in the seven-level rating scale was separately analysed for each subscale of MBI-SS using a rating scale model.⁴⁴ Linacre⁵³ created a set of seven guidelines for assessing the quality of rating scales (i.e. *directionality, monotonicity, category usage, distribution of ratings, rating scale fit, category coefficient order, and category coefficient locations*). Each of these rating scale guidelines will be considered in relation to the three MBI-SS subscales based on the RMM output.

For thresholds and category fit, one visual method of inspecting the distinction between thresholds is to examine the probability curves. Each category should have a distinct peak in the probability curve graph, illustrating that each in indeed the most probable response category for some portions of the measure variable.⁴⁴

Table 2
Rasch Measurement Model (RMM) item data for the three Maslach Burnout Inventory-Student Survey (MBI-SS) subscales ($N = 225$).

Parameter	RMM requirements	MBI-SS Exhaustion subscale	MBI-SS Cynicism subscale	MBI-SS Professional / Academic Efficacy subscale
Disordered thresholds	No	No	No	No
Number misfitting items	0	0	0	0
MBI-SS misfitting items		None	None	None
Item Model Infit <i>MnSq</i> Range Extremes	0.71 – 1.40	0.72 – 1.24	0.83 – 1.25	0.84 – 1.32
Item Model Infit <i>ZStd</i> Range Extremes	–2.00 to 2.00	–3.30 to 2.40	–1.60 to 0.00	–1.70 to 2.90
Item Model Outfit <i>MnSq</i> Range extremes	0.71 – 1.40	0.72 – 1.24	0.84 – 1.19	0.80 – 1.33
Item Model Outfit <i>ZStd</i> Range Extremes	–2.00 to 2.00	–3.30 to 1.50	–1.60 to 1.60	–2.20 to 3.20
Item reliability	> 0.80	0.98	0.95	0.98
Number Separate Item Strata	> 3.00	10.79	6.29	8.67
Item mean (<i>SD</i>) logits	0.00	0.00 (.68)	0.00 (.43)	0.00 (.59)
Variance accounted for by 1st factor	> 60%	82.1%	76.3%	66%
PCA (eigenvalue for 1st contrast)	≤ 2.00	6.30	13.60	9.00
Unexplained variance in contrasts 1–5 of PCA of residuals	< 5.00%	0.00%	0.00%	4.90%
Ceiling effect (% maximum extreme scores)	< 5.00%	0.89%	0.00%	0.44%
Floor effect (% minimum extreme scores)	< 5.00%	0.89%	18.67%	0.00%

Note: RMM = Rasch Measurement Model; MBI-SS = Maslach Burnout Inventory-Student Survey; Item spread is defined as the difference between maximum item logit score and minimum item logit score; *MnSq* = Mean Square; *ZStd* = standardized z; *SD* = Standard Deviation; PCA = Principle Components Analysis.

Table 3
Rasch Measurement Model (RMM) person data for the three Maslach Burnout Inventory-Student Survey (MBI-SS) subscales ($N = 225$).

Parameter	RMM requirements	MBI-SS Exhaustion subscale	MBI-SS Cynicism subscale	MBI-SS Professional / Academic Efficacy subscale
Number misfitting persons	< 5.00%	2.67%	5.33%	4.00%
Person spread		9.37	7.19	7.93
Person mean (<i>SD</i>) logits		0.73 (1.89)	–1.61 (1.66)	1.80 (1.52)
Person reliability	> 0.80	0.88	0.78	0.81
Person separation index	> 2.00	2.71	1.91	2.07
Number Separate Person Strata	> 3.00	3.95	2.88	3.09
Difference between person and items means	< 1.00	0.73	–1.61	1.80
DIF by gender (Item Number [DIF contrast])	> 0.50 logits, $p < .05$	0.00	0.00	0.00
MBI-SS items exhibiting DIF by gender		None	None	None

Note: RMM = Rasch Measurement Model; MBI-SS = Maslach Burnout Inventory-Student Survey; *SD* = Standard Deviation; Person spread is defined as the difference between maximum person logit score and minimum person logit score; DIF = Differential Item Functioning.

3.3. Reliability

The reliability of the MBI-SS items is evaluated in the form of the corrected item-total correlation (item reliability) to assess the homogeneity of the items in each subscale (considered adequate if it is > 0.80). Person reliability indices include the person reliability coefficient, interpreted as the replicability of the person ordering and the person separation index, which identifies the number of statistically distinct ability strata of the individuals in the sample.⁴⁴ A minimum of

two distinct ability strata are required according to Boone et al.⁵⁵

3.4. Differential Item Functioning (DIF)

DIF was examined by considering the difference between logit scores for each of the three MBI-SS subscale items based on gender. In this instance, logit values of the scale items based on first on gender (males versus females) were generated and examined for potential significant differences using *t*-test compar-

Table 4
RMM logit measures and fit statistics for the three Maslach Burnout Inventory-Student Survey (MBI-SS) subscales ($N = 225$).

MBI-SS Subscale items	RMM Logit item Measure	Logit item Measure SE	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Point Measure Correlation
Exhaustion Subscale							
Item 4	0.95	0.08	0.89	-1.20	0.87	-1.50	.85
Item 5	0.59	0.08	0.72	-3.30	0.72	-3.30	.87
Item 1	-0.11	0.08	1.03	0.40	1.04	0.40	.80
Item 3	-0.61	0.08	1.24	2.40	1.15	1.50	.80
Item 2	-0.82	0.09	1.13	1.30	1.05	0.50	.79
Cynicism Subscale							
Item 9	0.65	0.10	1.25	2.10	1.19	1.60	.76
Item 8	0.03	0.09	1.00	0.00	0.96	-0.30	.83
Item 6	-0.12	0.09	0.95	-0.40	0.92	-0.70	.84
Item 7	-0.56	0.09	0.83	-1.60	0.84	-1.60	.87
Professional / Academic Efficacy Subscale							
Item 10	0.66	0.08	1.32	2.90	1.33	3.20	.68
Item 15	0.50	0.08	0.86	-1.40	0.80	-2.20	.76
Item 11	0.44	0.09	0.99	-0.10	0.96	-0.40	.77
Item 12	-0.03	0.09	0.85	-1.60	0.81	-2.10	.80
Item 14	-0.70	0.10	0.84	-1.70	0.83	-1.80	.74
Item 13	-0.86	0.10	1.20	1.90	1.15	1.50	.71

Note: RMM = Rash Measurement Model; SE = Standard Error; MnSq = Mean Square; ZStd = standardized z .

isons. If any significant differences were found between the two sets of RMM logit scores, then DIF would be present.^{44,55–57}

4. Results

4.1. Demographic data

The total sample consisted of 200 female (88.9%) and 25 male (11.1%) students with an average age of 21.64 years ($SD = 4.11$). Other participant group demographic variables are reported in Table 1.

4.2. Scale unidimensionality

The principal component analysis (PCA) of the RMM residuals did not indicate any significant additional dimensions in the item residuals (see Tables 2 and 3). For the Exhaustion subscale, the initial PCA revealed that 82.1% of the total variance was explained by the dimension and 6.3% of the variance was unexplained in the first contrast. For Cynicism and Professional / Academic Efficacy, the initial PCA results were 76.3% and 66% of the total variance was accounted for respectively.

No ceiling effect was observed for the three subscales because less than 1% of the participants had difficulties completing the three subscales ($< 5\%$ of the total sample) and similarly, no floor effect was found for these subscales, except to Cynicism with more than 5% of sample achieving minimum scores.⁵⁸

4.3. Item and person fit statistics

All of the MBI-SS items were included in the Winsteps RMM calibration. The RMM analysis results of the three subscales of MBI-SS are reported in Tables 2–4 respectively. In the goodness-of-fit analysis, all items from each of the three MBI-SS subscales met the fit statistics requirements; that is, all the items exhibited acceptable fit statistics for the RMM (see Tables 2 and 3). The MnSq infit statistics varied from 0.72 to 1.24, 0.83 to 1.25, and 0.84 to 1.32 for the MBI-SS Exhaustion, Cynicism, and Professional / Academic Efficacy subscales respectively. The ZStd infit statistics ranged from 2.40 to -3.30 for Exhaustion, 0.00 to -1.60 for Cynicism and 2.90 to -1.70 for Professional / Academic Efficacy. These results indicated acceptable levels of scalability and unidimensionality for the Cynicism subscale, and closely acceptable levels

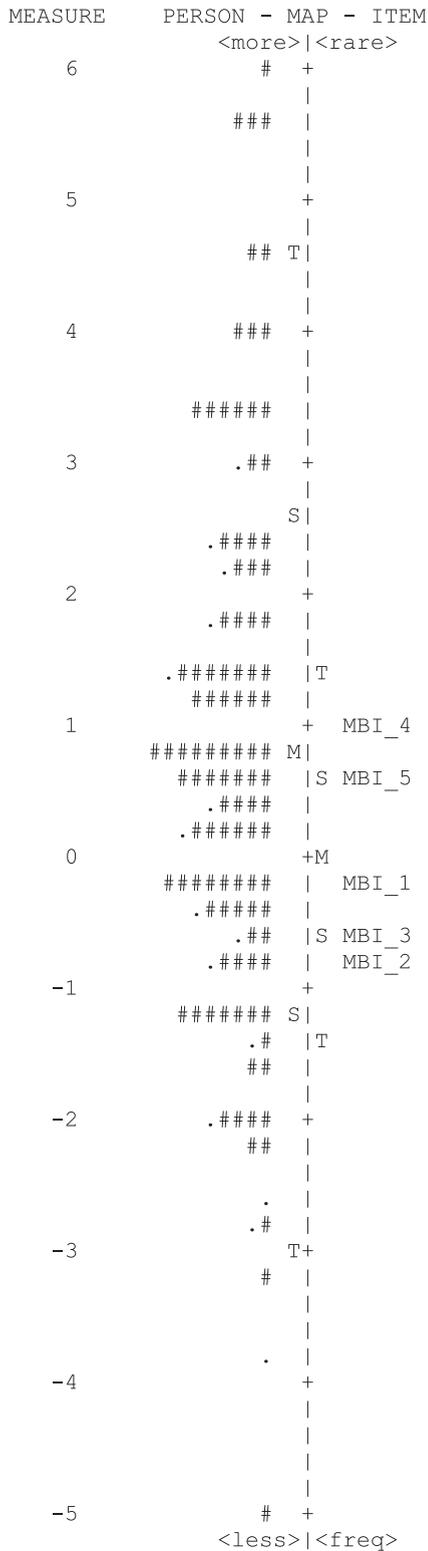


Fig. 1. Wright Person-Item Rasch map for the MBI-SS Exhaustion subscale ($N = 225$). Note: each '#' represents two participants and each '.' represents one participant.

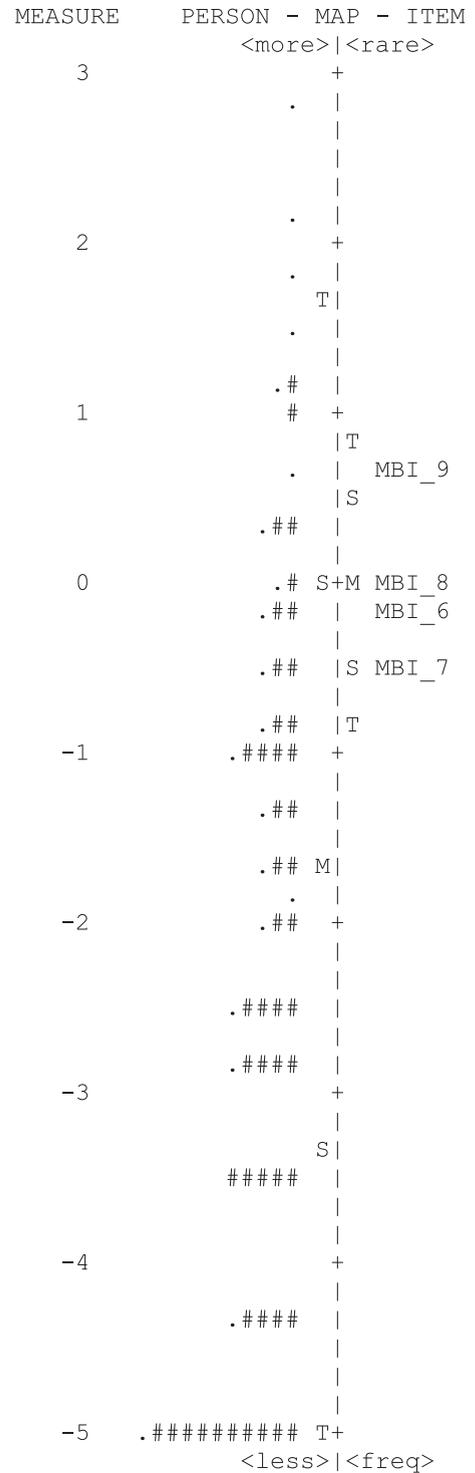


Fig. 2. Wright Person-Item Rasch map for the MBI-SS Cynicism subscale ($N = 225$). Note: each '.' represents four participants and each '#' represents one to three participant(s).

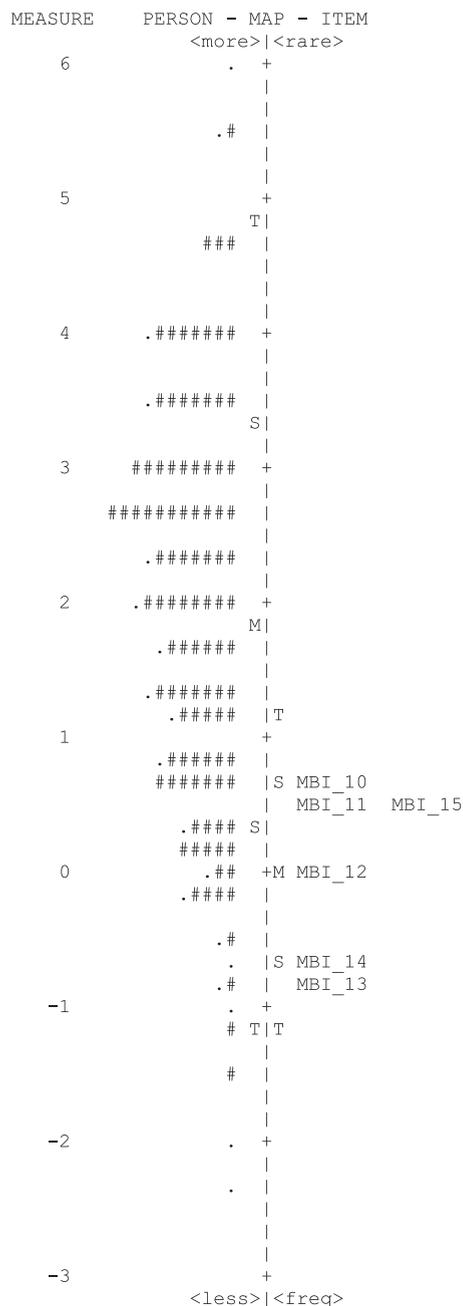


Fig. 3. Wright Person-Item Rasch map for the MBI-SS Professional / Academic Efficacy subscale (N = 225). Note: each “.” represents two participants and each “.” represents one participant.

for the Exhaustion and Professional / Academic Efficacy subscales.

The Wright Person-Item Rasch maps⁴⁴ for the three MBI-SS Efficacy subscales are located in Figs. 1–3 respectively. The Wright Person-Item maps chart the person ability logit scores against the item difficulty logit scores.^{44,55} It provides a visual representation of

how the difficulty levels of the three MBI-SS Efficacy subscale items match the ability levels of the participant group in a hierarchical representation.

In the Rasch goodness-of-fit person-response analyses, the results indicated that only the 2.67% of the sample misfit on the Exhaustion, 5.33% on the Cynicism, and 4% on the Professional / Academic Efficacy subscales respectively. The misfitting persons could be consequence of varying levels of attention by students when filling out the MBI-subscale or being stressed when responding to the subscale items.

4.4. Scalability/interval level scaling

The results indicated that the seven-level rating scale used in the three MBI-SS subscales fit appropriately to the predictive RMM.⁴⁴ Table 5 summarizes the rater use of the rating scale for the sample of undergraduate occupational therapy students. The average measure observed, average measure expected, rating category usage, outfit MSE, category measure and Andrich threshold are reported for each of the seven rating scale categories for each of the three MBI-SS subscales. Inspection of the rating scale structure of the three MBI-SS subscales against the seven guidelines listed by Linacre⁵³ in Table 6 indicates that the rating scale categories (i.e. 0 → 6) for the three subscales are for the most part generating meaningful measures of students’ levels of burnout.

Rasch rating scale parameters are also referred to as Rasch-Andrich thresholds and relate to category probabilities. “The Rasch-Andrich threshold is defined as the difficulty of observing category k relative to category $k - 1$. It can be seen as the points of which categories k and $k - 1$ are equally likely to be observed in the continuum. There is a modal perspective on the category boundaries on the latent variable if the categories are visible and ordered”.⁵⁹

The results indicated that the seven-level rating scale used in the three MBI-SS subscales fit appropriately to the predictive Rasch Model (as depicted in Figs. 4 and 5 respectively). The probably curves for the rating scales of the three subscales illustrated that each threshold estimate represented a separate point on the measure variable and each response category was the most probable category for some part of the continuum.³⁰

In relation to *directionality*, the rating categories appear to be aligned with the latent variable, as indicated by the relatively close match between observed and expected average measures for the three MBI-SS subscales in Table 5.

Table 5
Rating Scale Structure of the three Maslach Burnout Inventory-Student Survey (MBI-SS) subscales ($N = 225$).

	Average Measure Observed	Average Measure Expected	Category Usage (%)	Outfit MSE	Category Measure	Andrich Threshold
Exhaustion subscale rating scale categories						
0	−2.60	−2.75	37 (3)	1.22	−3.53	None
1	−1.87	−1.78	100 (9)	0.83	−2.32	−1.92
2	−0.81	−0.84	159 (14)	0.98	−1.46	−2.14
3	0.03	0.06	246 (22)	0.81	−0.47	−1.25
4	1.01	0.99	226 (20)	0.89	0.84	0.18
5	2.29	2.24	249 (22)	1.07	2.64	1.36
6	4.02	4.12	107 (10)	1.12	4.92	3.76
Cynicism subscale rating scale categories						
0	−3.37	−3.43	330 (37)	1.10	−4.33	None
1	−2.47	−2.31	203 (23)	0.97	−2.41	−3.10
2	−1.11	−1.22	152 (17)	0.67	−1.03	−1.46
3	−0.28	−0.35	110 (12)	0.67	0.07	−0.45
4	0.53	0.47	71 (8)	1.01	1.21	0.49
6 ^a	1.11	1.94	9 (1)	2.57	3.90	2.48
5 ^a	1.31	1.26	22 (2)	1.39	2.40	2.03
Professional / Academic Efficacy subscale rating scale categories						
0	−1.75	−1.93	11 (1)	1.11	−3.53	None
1	−0.97	−1.19	16 (1)	1.66	−2.32	−1.92
2	−0.51	−0.45	60 (4)	1.12	−1.46	−2.14
3	0.29	0.29	191 (14)	0.97	−0.47	−1.25
4	1.13	1.21	331 (25)	0.74	0.84	0.18
5	2.44	2.37	500 (37)	0.94	2.64	1.36
6	3.58	3.62	239 (18)	1.09	4.92	3.76

^aThe rating scale categories 5 and 6 for the Cynicism subscale are disordered.

The monotonic increase in category average measures of the three MBI-SS subscales' rating categories demonstrates adherence to second guideline of *monotonicity*. This index indicates agreement between the observed and intended ordering of rating scale categories, except for Cynicism since the average measures in the rating scale in this subscale were disordered for items 5 and 6.

The *category usage* and *distribution of ratings* guidelines indicate that in general the category usage frequencies and percentages for the three MBI-SS subscales demonstrate a reasonable coverage of ratings across scale categories. Linacre⁵³ has indicated that as a rule of thumb that rating categories with fewer than ten participant observations may reduce the precision and stability of these estimates. It should be noted that the 6th rating category of the Cynicism subscale only had nine category usages whereas the 0 rating category of the Efficacy subscale had 11 category usages. This indicates that the rating categories for these two

subscales should be investigated further with the possible elimination or collapsing of these rating scale categories.

In relation to the *distribution of ratings* of the seven rating categories of the three MBI-SS subscales, several had a low percentage of usage. These categories included the rating category 0 of the Exhaustion subscale, rating categories 5 and 6 for the Cynicism subscale, and rating categories 1, 2, and 3 of the Professional / Academic Efficacy subscale (see Tables 5 and 6).

Then next guideline is *rating scale fit* where mean square error (MSE) statistics for residual analyses provide an indication of model-data fit. Linacre⁵³ suggests using the outfit MSE statistics in this instance since they are sensitive to outliers. The expected value of the outfit MSE is 1.00. High outfit MSE indicates a high degree of randomness and the use of rating categories in unexpected ways. Low outfit MSE results indicate rating patterns that are more uniform than

Table 6

Linacre's rating scale quality assessment guideline requirements applied to the three Maslach Burnout Inventory-Student Survey (MBI-SS) subscales ($N = 225$).

Rating Scale Quality Requirements	Exhaustion Subscale	Cynicism Subscale	Professional / Academic Efficacy Subscale
Directionality	✓	✓	✓
Monotonicity	✓	✓	✓
Category Usage	✓ for rating categories 0 → 6	× for rating category 6 ✓ for rating categories 0 → 5	× for rating category 0 ✓ for rating categories 1 → 6
Uniform Distribution of Ratings	× for rating category 0 ✓ for rating categories 1 → 6	× for rating categories 5 & 6 ✓ for rating categories 0 → 4	× for rating categories 0 → 2 ✓ for rating categories 3 → 6
Rating Scale Fit	× for rating categories 0 & 3 ✓ for rating categories 1, 2, 4, 5, & 6	× for rating categories 2, 3, 5 & 6 ✓ for rating categories 0, 1, & 4	× for rating categories 1 & 4 ✓ for rating categories 0, 2, 3, 5, & 6
Category Coefficient Order	✓	✓	✓
Category Coefficient Locations	✓	✓	✓

✓ denotes that the rating scale quality assessment guideline requirement has been met whereas × infers that the rating scale quality assessment guideline requirement has not been met.

expected. For the three MBI-SS subscales, the majority of the rating scale categories were near the expected value of 1.00. However, the Cynicism subscale had two rating categories (2 and 3) that were below 1.00 and two rating categories (5 and 6) that were above the 1.00 level. Therefore, this finding indicates that Cynicism subscale had a lower level of *rating scale fit* compared to the other two MBI-SS subscales.

The next guideline is *category coefficient order* where rating category measures need to reflect the planned order of categories in terms of progression. Evidence of this index is the monotonic increase of the rating category coefficients for all three MBI-SS subscales as reported in Tables 5 and 6. The final guideline is *category coefficient locations*. Linacre⁵³ described a range of threshold location differences between 1.40 and 5.00 logits as a strong indicator that rating scale categories are unique. The differences between the rating scale categories for the three MBI-SS rating scales indicate that they are distinctive (Figs. 4–6).

4.5. Reliability

Person results also fit the RMM requirements, where the mean person measure was 0.73 ($SD = 1.89$), -1.61 ($SD = 1.66$), and 1.80 ($SD = 1.52$) for Exhaustion, Cynicism, and Professional / Academic Efficacy respectively. The person reliability was higher than 0.80 for two of the subscales, except for the

Cynicism subscale with a result of 0.78. The person separation index results were higher than 2.00 logits for the three subscales, except for the Cynicism subscale. These results are reported in Tables 2 and 3.

4.6. Differential Item Functioning (DIF)

None of the items from the three MBI-SS subscales exhibited DIF based on gender. However, the DIF analyses may present a potential bias due to the unequal distribution between the two groups being compared. In other words, because there were many more female respondents ($n = 200$) than male respondents ($n = 25$), this distribution may have biased the DIF results.

5. Discussion

The aim of this study was to investigate the unidimensionality, scalability, DIF, rating scale structure, and reliability of the three individual MBI-SS subscales using the RMM. The unidimensionality for each subscale of the MBI-SS was verified. All the items from each of the three MBI-SS subscales exhibited acceptable or closely acceptable fit statistics for the RMM. Only a small percentage of the items in the sample were misfitting when completed by the respondents. The results indicated that the seven-level rating scale used in the three MBI-SS subscales also fit appropriately to the RMM requirements. Item and person reliability analyses showed that the MBI-SS

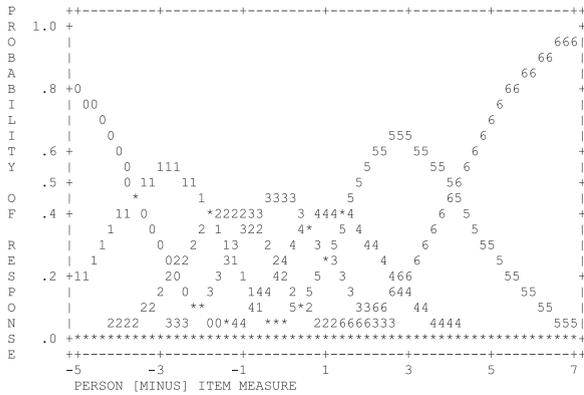


Fig. 4. Rating scale structure and functioning of the MBI-SS Exhaustion subscale (N = 225).

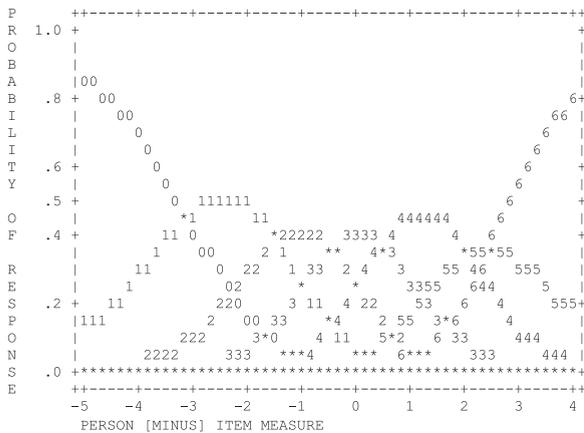


Fig. 5. Rating scale structure and functioning of the MBI-SS Cynicism subscale (N = 225).

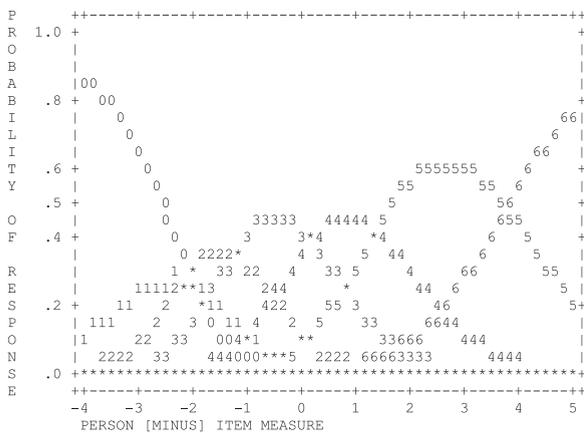


Fig. 6. Rating scale structure and functioning of the MBI-SS Professional / Academic Efficacy subscale (N = 225).

items provided precise or closely precise estimates defining the three components of student burnout. None of the MBI-SS subscale items exhibited DIF based on gender, despite the two comparison groups being unequal in size.

The findings indicated that the MBI-SS is a useful scale for use in university education contexts. This scale could assist in identifying specific occupational therapy undergraduate students who might be feeling exhausted as a consequence of the demands of their academic studies or other life demands such as external work commitments or family responsibilities. In addition, the MBI-SS could be employed to detect when students might be experiencing cynical attitudes, poor engagement in their academic studies or not feeling competent in their student role. Students would not be learning effectively if they were experiencing indicators of burnout.^{18,20,60–62} Hence, early detection and monitoring of students’ burnout would assist academic and fieldwork personnel in providing targeted intervention or remediation strategies to combat students’ burnout. This information could be utilized when planning new strategies and approaches to motivate students. The MBI-SS could also be used to evaluate the impact or effectiveness of programs put in place to assist with students managing the stressors associated with academic pursuits.

5.1. Scale unidimensionality

In the current study, the RMM results indicated that the MBI-SS items confirmed the unidimensionality of the three sub-constructs of Exhaustion, Cynicism, and Professional / Academic Efficacy as part of the whole construct of burnout. For this reason, the importance of the unidimensional structure of each of the three individual MBI-SS subscales in defining the overall construct referred to as ‘burnout syndrome’ was highlighted.

The first English language version of MBI-SS was used by Schaufeli et al.²⁸ after recruiting participant groups from three European countries: Spain, Portugal, and the Netherlands. In their study, using a Classical Test Theory approach, factor analysis results indicated that a three factors model fitted the sample data. Other studies findings using a confirmatory factor approach also corroborated the hypothesized three-factor model of the MBI-SS.^{27,36,39,63–66} Hence, the previously completed studies have confirmed that each one of the MBI-SS subscales measures a different sub-dimension of the overall burnout syndrome construct.

However, in one study using a back-translated Japanese version of the MBI-SS that also used a confirmatory factor analysis approach, Tsubakita and Shimazaki found that “the hypothesized three factor model did not meet the respective criteria” and “subsequent analysis revealed that a bi-factor model fit the data better than did the hypothesized or modified three factor models”.²⁹ However, it should be noted that the different language version of the MBI-SS and the Classical Test Theory approach used to analyze its factor structure could have been contributing reasons why the three-factor structure was not supported in Tsubakita and Shimazaki²⁹ study.

A number of other studies have been completed that have investigated the factor structure of the MBI-SS using Classical Test Theory-based analysis including exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and principal components analysis (PCA) with varying results. Yavuz and Dogan³² examined the factorial validity of the MBI-SS in a sample of 1020 Turkish high school students. “Velicer’s MAP Test, Parallel analysis and Confirmatory and Explanatory factor analysis were used to determine the construct validity of scale.”³² According to Yavuz and Dogan³² findings, the MAP Test, Parallel analysis and EFA did not support the three dimensional structure of the MBI-SS whereas the CFA results did endorse the original factor structure reported by Schaufeli et al.^{28,32}

Simancas-Pallares et al.³⁵ investigated the validity of the MBI-SS in a sample of 886 dental students from Cartagena, Columbia. Using EFA and CFA, the factor structure of the MBI-SS did not fit the original proposed dimensional structure. Ilic et al.²⁷ evaluated the factorial validity of the MBI-SS involving a sample of 760 Serbian medical students. “Principal Component Analysis with Oblimin rotation indicated 3 main components that explained 64.9% variance.”²⁷ while CFA generated good fit indices. Finally, in a study involving 667 French students, Fay-Dumanget et al.³⁴ determined that the three-dimensional structure of the MBI-SS was substantiated.

5.2. Scalability

Interval level scaling of the Exhaustion, Cynicism, and Professional / Academic Efficacy subscale items was confirmed by the logit scores obtained. All items had infit and outfit statistics within the acceptable range or very close to these, indicating that the MBI-SS sub-dimensions fit the unidimensionality expectations of the RMM. In other words, each subscale of the MBI-SS measured a unidimensional part of the whole construct

of burnout syndrome in Australian undergraduate occupational therapy students. Other studies have evaluated the measurement properties of the MBI-SS involving undergraduate students from different countries. Although these investigations used a Classical Test Theory approach, they have reported similar results.^{63–67}

5.3. Reliability

The item reliability was explored by using output from the Winsteps Rasch analysis program. These indices provided precise estimates of item consistency in the sample of Australian university students. Results obtained in this study were similar to those reported in other studies.^{50,64,66,67} In the pilot study conducted by Campos et al.⁶⁴ the MBI-SS’s reliability was investigated using the Interclass Correlation Coefficient and it was concluded that the reproducibility of the dimensions was good for the Exhaustion and Cynicism subscales, and was moderate for the Professional / Academic Efficacy subscale.⁶⁴

As part of the RMM output, person reliability indices included the person reliability coefficient and the person separation index. The first one demonstrated a good reliability related to the replicability of the person ordering. The second one showed that the number of statistically distinct ability strata of the individuals in the sample in each subscale were higher than two, in each of the MBI-SS Exhaustion and Professional / Academic Efficacy subscales. This finding indicates the number of ability levels that can be distinguished by the scale. Then, this scale generated at least two-person strata which can distinguish the sample into two different abilities levels to complete the scale. However, although the Cynicism subscale exhibited less than two distinct ability strata, the results were very close to the limit of two strata. To obtain less than two-person strata levels would mean that the MBI-SS can only distinguish one or two ability levels to complete this scale which would be problematic. However, this situation was not the case for the three MBI-SS subscales.

In the refereed literature, the internal consistency in the form of Cronbach’s α coefficients of the three MBI-SS scales have been reported previously.^{27,29,32–37} Although the Cronbach’s α coefficients are based on Classical Test Theory principles, it still provides a point of comparison for the person reliability and item reliability indices generated by the RMM Winsteps program. Ilic²⁷ reported that the “overall Cronbach’s α coefficient of the MBI-SS questionnaire was 0.757,

while the Cronbach's α coefficients for Emotional Exhaustion, Cynicism, and [Professional /]Academic Efficacy were 0.869, 0.856, and 0.852, respectively."²⁷ Yavuz and Dogan³² reported that the Cronbach's α coefficients for MBI-SS Emotional Exhaustion, Cynicism, and Professional / Academic Efficacy subscales were 0.838, 0.844, and 0.875 whereas Tsubakita and Shimazaki²⁹ reported Cronbach's α coefficients of 0.76, 0.85, and 0.78, respectively.

5.4. Differential item functioning

The findings of the present study indicated that MBI-SS items did not exhibit DIF since all items presented stable item difficulty across gender, that is, male and female participants did not respond to these items differently. Similarly, other studies indicate that the MBI-SS items also operate in a gender-neutral manner when the MBI-SS scale is completed by respondents.^{26,53} The lack of DIF based on gender should be considered a strength of the three MBI-SS subscales.

5.5. Limitations

Four limitations were inherent in this study. Firstly, participants took part in the study on a voluntary basis and were not randomly selected. It was not considered ethical to mandate that participants take part in the study. Secondly, given the MBI-SS is a self-report scale, participants may have answered items in a socially desirable manner, hence respondent bias may have been present. Thirdly, students were recruited from one specific occupational therapy program located in one geographical location, therefore the generalizability of the results may be limited. However, when local independence was examined in the RMM Win-steps program, it was supported. Finally, since the sample had an unequal distribution across the two gender groups (25 men versus 200 women), there may not be sufficient statistical power to identify DIF between these groups. Therefore, the DIF results should be interpreted with caution.

5.6. Future research

It is recommended that a similar study be carried out with other student groups such as physiotherapy, social work, nursing, education, engineering and business. It is suggested that future sample groups be randomly selected. The cross-cultural validity of the three factors related to burnout as measured by the MBI-SS could be investigated by recruiting undergraduate occupational

therapy students from education programs located in different countries. Finally, the dimensionality, DIF, rating scale structure and reliability of the MBI-SS could be examined by recruiting a group of occupational therapy students enrolled in a graduate-entry masters or entry-to-practice clinical doctorate course.

6. Conclusion

The study findings indicated that the three individual subscales of the Maslach Burnout Inventory-Student Survey can be used to assess the burnout syndrome in undergraduate students enrolled in occupational therapy courses. The three-dimensional structure of the MBI-SS was consistent with previous studies when completed by a group of occupational therapy undergraduate students. The scalability, structural validity and reliability of the MBI-SS fitted with RMM requirements. The appropriateness of the rating scale was upheld for the three individual MBI-SS subscales across individual items using its seven-level rating scale.

Acknowledgements

The undergraduate Occupational Therapy students from Monash University who volunteered to take part in this study are thanked for their time and input.

Author contributions

JM and TB contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript.

Disclosure

None.

Ethical approval

Ethical approval was granted by Monash University Human Research Ethics Committee approval on the 9 April 2015 (Ethics approval ID: 2015-6069-5898).

Declaration of interest

None.

Declaration and verification

The work described has not been published previously, it is not under consideration for publication elsewhere, its publication is approved by all authors and tacitly by the responsible authorities where the work was carried out, and if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder.

All authors contributed substantially to the development, data collection, analysis of results, and write-up this manuscript and are in agreement regarding its contents.

Funding

José Manuel Pérez-Mármol was supported by “Becas del Plan Propio” from the University of Granada (Spain).

References

- Schaufeli W, Enzmann D. *The Burnout Companion to Study and Practice: A Critical Analysis*. Philadelphia, IL: CRC Press; 1998.
- Maslach C, Jackson SE. The measurement of experienced burnout. *J Organ Behav* 1981;2(2):99–113 <http://dx.doi.org/10.1002/job.4030020205>.
- Leung SS, Wah-Mak Y, Yu-Chui, Y, et al. Occupational stress, mental health status and stress management behaviors among secondary school teachers in Hong Kong. *Health Educ J* 2009;68(4):328–343 <http://dx.doi.org/10.1177/0017896909349255>.
- Tomljenovic M, Kolaric B, Stajduhar, D, et al. Stress, depression and burnout among hospital physicians in Rijeka, Croatia. *Psychiatr Danub* 2014;26:450–458.
- Bassett H, Lloyd C. Occupational therapy in mental health: managing stress and burnout. *Br J Occup Ther* 2001;64(8):406–411 <http://dx.doi.org/10.1177/030802260106400807>.
- Cahalan E, Manigandan C. Levels of stress and burnout in occupational therapists working in mental health in Ireland. *Ir J Occup Ther* 2010;2011;38(2):13–24 <http://dx.doi.org/10.1177/2158244015604694>.
- Edwards H, Durette D. The relationship between professional identity and burnout among occupational therapists. *Occup Ther Health Care* 2010;24(2):119–129 <http://dx.doi.org/10.3109/07380570903329610>.
- Gupta S, Paterson ML, Lysaght, RM, et al. Experiences of burnout and coping strategies utilized by occupational therapist. *Can J Occup Ther* 2012;79(2):86–95 <http://dx.doi.org/10.2182/cjot.2012.79.2.4>.
- Lee SH, Bang YS. The effect of emotional labor and burnout on the job attitude of occupational therapist. *J Korean Soc Occup Ther* 2015;23:95–108 <http://dx.doi.org/10.14519/jksot.2015.23.1.08>.
- Lloyd C, McKenna K, King R. Sources of stress experienced by occupational therapists and social workers in mental health settings. *Occup Ther Int* 2005;12(2):81–94 <http://dx.doi.org/10.1002/oti.17>.
- Painter J, Akroyd D, Elliot, S, et al. Burnout among occupational therapists. *Occup Ther Health Care* 2003;17(1):63–78 http://dx.doi.org/10.1080/J003v17n01_06.
- Poulsen AA, Meredith P, Khan, A, et al. Burnout and work engagement in occupational therapists. *Br J Occup Ther* 2014;77(3):156–164 <http://dx.doi.org/10.4276/030802214X13941036266621>.
- Rogers JC, Dodson SC. Burnout in occupational therapists. *Am J Occup Ther* 1988;42(12):787–792.
- Sweeney G, Nichols K. Stress experiences of occupational therapists in mental health practice arenas: a review of the literature. *Int J Soc Psychiatry* 1996;42(2):132–140 <http://dx.doi.org/10.1177/002076409604200207>.
- Wolf K. Profession-specific stress factors for occupational, physio, and speech therapists. *Ergother Rehabil* 2011;50(12):22–28.
- Yang YA, Hur JG, Noh YM, Lee GC. The study of job stress in occupational therapist. *J Ergon Soc Korea* 2007;26(3):1–9 <http://dx.doi.org/10.5143/JESK.2007.26.3.001>.
- Zeman EA, Harvison N. Burnout, stress, and compassion fatigue in occupational therapy practice and education: a call for mindful, self-care protocols. *Perspect: Expert Voices Health Health Care Natl Acad Med* 2017. Retrieved from: (<https://nam.edu/wp-content/uploads/2017/03/Burnout-Stress-and-Compassion-Fatigue-in-Occupational-Therapy-Practice-and-Education-A-Call-for-Mindful-Self-Care-Protocols.pdf>).
- Backović DV, Ilić Živojinović J, Maksimović, J, et al. Gender differences in academic stress and burnout among medical students in final years of education. *Psychiatr Danub* 2012;24(2):175–181.
- Robin TG, Roberts RM, Sarris A. Burnout and engagement in health professional students: the relationships between study demands, study resources and personal resources. *Australas J Organ Psychol* 2015;8(e1):1–13 <http://dx.doi.org/10.1017/orp.2014.7>.
- Schaufeli WB, Martinez IM, Pinto, AM, et al. Burnout and engagement in university students: a cross-national study. *J Cross Cult Psychol* 2002;33(5):446–481 <http://dx.doi.org/10.1177/0022022102033005003>.
- Pfeifer TA, Kranz PL, Scoggin AE. Perceived stress in occupational therapy students. *Occup Ther Int* 2006;15(4):176–192 <http://dx.doi.org/10.1002/oti.256>.
- Tyrrell J, Smith H. Levels of psychological distress among occupational therapy students. *Br J Occup Ther* 2015;59(8):365–371 <http://dx.doi.org/10.1177/030802269605900805>.
- Everly JS, Poff DW, Lampert, N, et al. Perceived stressors and coping strategies of occupational therapy students. *Am J Occup Ther* 1994;48(11):1022–1028 <http://dx.doi.org/10.5014/ajot.48.11.1022>.
- Dias M, Vellarde GC, Olej, B, et al. Effects of electroacupuncture on stress-related symptoms in medical students: a randomised placebo-controlled study. *Acupunct Med* 2014;32(1):4–11 <http://dx.doi.org/10.1136/acupmed-2013-010408>.
- Fang DZ, Young CB, Golshan, S, et al. Burnout in premedical undergraduate students. *Acad Psychiatry* 2012;36(1):11–16 <http://dx.doi.org/10.1176/appi.ap.10080125>.
- Galán F, Ríos-Santos JV, Polo-Padillo, J, et al. Burnout, depression and suicidal ideation in dental students. *Med Oral Patol Oral Cir Bucal* 2014;19(3):1–6 <http://dx.doi.org/10.4317/medoral.19281>.
- Ilic M, Todorovic Z, Jovanovic, M, et al. Burnout syndrome among medical students at one University in Serbia: validity and reliability of the Maslach Burnout Inventory-Student Survey. *Beh Med* 2017;43(4):323–328 <http://dx.doi.org/10.1080/08964289.2016.1170662>.
- Schaufeli WB, Salanova M, González-Romá, V, et al. The measurement of engagement and burnout: a two-sample

- confirmatory factor analytic approach. *J Happiness Stud* 2002;3(1):71–92 <http://dx.doi.org/10.1023/A:1015630930326>.
29. Tsubakita T, Shimazaki K. Constructing the Japanese version of the Maslach Burnout Inventory-Student Survey: confirmatory factor analysis. *Jpn J Nurs Sci* 2016;13(1):183–188 <http://dx.doi.org/10.1111/jjns.12082>.
 30. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 2014.
 31. Streiner D, Norman G, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*, 5th ed., Oxford, UK: Oxford University Press; 2015.
 32. Yavuz G, Dogan N. Maslach Burnout Inventory-Student Survey (MBI-SS): a validity study. *Proc - Social Behav Sci* 2014;116:2453–2457 <http://dx.doi.org/10.1016/j.sbspro.2014.01.590>.
 33. Rostami Z, Abedi MR, Schaufeli, WB, et al. The psychometric characteristics of Maslach Burnout Inventory Student Survey: a study students of Isfahan University. *Zahedan J Res Med Sci* 2014;16(9):55–58. (https://zjrms.ir/browse.php?a_id=2172&sid=1&slc_lang=en).
 34. Faye-Dumanget C, Carré J, Le Borgne, M, et al. French validation of the Maslach Burnout Inventory-Student Survey (MBI-SS). *J Eval Clin Pract* 2017;23:1247–1251 <http://dx.doi.org/10.1111/jep.12771>.
 35. Simancas-Pallares MA, Fortich Mesa N, González Martínez FD. Validez y consistencia interna del Inventario Maslach para burnout en estudiantes de Odontología de Cartagena, Colombia. *Rev Colomb Psiquiat* 2017;46:103–109 <http://dx.doi.org/10.1016/j.rcpeng.2017.05.009>.
 36. Gumz A, Erices R, Brähler, E, et al. Factorial structure and psychometric criteria of the German translation of the Maslach Burnout Inventory – Student Version by Schaufeli et al. (MBI-SS). *Psychother Psych Med* 2013;63(02):77–84 <http://dx.doi.org/10.1055/s-0032-1323695>.
 37. Hederich-Martinez C, Caballero-Dominguez CC. Validation of Maslach Burnout Inventory-Student Survey (MBI-SS) in Colombian academic context. *Rev CES Psicol* 2016;9(1):1–15. (<http://revistas.ces.edu.co/index.php/psicologia/article/view/3511/2508>).
 38. Shin H, Puig A, Lee, J, et al. Cultural validation of the Maslach Burnout Inventory for Korean Students. *Asia Pac Educ Rev* 2011;12(4):633–639 <http://dx.doi.org/10.1007/s12564-011-9164-y>.
 39. Galan F, Sanmartín A, Polo, J, et al. Burnout risk in medical students in Spain using the Maslach Burnout Inventory-Student Survey. *Int Arch Occup Environ Health* 2011;84(4):453–459 <http://dx.doi.org/10.1007/s00420-011-0623-x>.
 40. Mokkink L, Terwee C, Patrick, D, et al. Protocol of the COSMIN study: consensus-based Standards for the selection of health Measurement Instruments. *BMC Med Res Method* 2006;6(1):2 <http://dx.doi.org/10.1186/1471-2288-6-2>.
 41. Mokkink L, Terwee C, Patrick, D, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63(7):737–745 <http://dx.doi.org/10.1016/j.jclinepi.2010.02.006>.
 42. Newton PE, Shaw SD. Disagreement over the best way to use the word 'validity' and options for reaching consensus. *Assess Educ: Princ Policy Pract* 2016;23(2):178–197 <http://dx.doi.org/10.1080/0969594X.2015.1037241>.
 43. Rios J, Wells C. Validity evidence based on internal structure. *Psicothema* 2014;26(1):108–116 <http://dx.doi.org/10.7334/psicothema2013.260>.
 44. Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, third edition, New York: Routledge / Taylor and Francis Group; 2015.
 45. Tennant A, Conaghan PG. 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 Care Res* 2007;57(8):1358–1362 <http://dx.doi.org/10.1002/art.23108>.
 46. Smith RM. *Introduction to Rasch Measurement: Theory, Models and Applications*. Maple Grove, MN: JAM Press; 2004.
 47. Karabatos G. The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *J Appl Meas* 2001;2(4):389–423.
 48. Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002;3(2):205–231.
 49. Linacre JM. *Winsteps® (Version 3.93.0) [Computer Software]*. Beaverton, Oregon: Winsteps.com; 2017. [Retrieved October 1, 2017]. Available from (<http://www.winsteps.com/>).
 50. Brown T, Rodger S. An evaluation of the validity of the Test of Visual Perceptual Skills-Revised (TVPS-R) using the Rasch Measurement Model. *Br J Occup Ther* 2009;72(2):65–78 <http://dx.doi.org/10.1177/030802260907200204>.
 51. Brown T, Unsworth C, Lyons C. An evaluation of the construct validity of the Developmental Test of Visual-Motor Integration using the Rasch Measurement Model. *Aust Occup Ther J* 2009;56(6):393–402 <http://dx.doi.org/10.1111/j.1440-1630.2009.00811.x>.
 52. Cadime I, Ribeiro I, Viana FL, Santos S, Prieto G. Calibration of a reading comprehension test for Portuguese students. *An Psicol* 2014;30(3):1025–1034 <http://dx.doi.org/10.6018/analesps.30.3.172611>.
 53. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas* 2002;3(1):85–106 <http://dx.doi.org/10.12691/ajphr-3-4-3>.
 54. Linacre JM. Investigating rating scale category utility. *J Outcome Meas* 1999;3(2):103–122.
 55. Boone WJ, Staver JR, Yale MS. *Rasch Analysis in the Human Sciences*. New York: Springer; 2013.
 56. Babiar TC. Exploring differential item functioning (DIF) with the Rasch model: a comparison of gender differences on eighth grade science items in the United States and Spain. *J Appl Meas* 2011;12(2):144–164.
 57. Wyse AE, Mapuranga R. Differential item functioning analysis using Rasch item information functions. *Intern J Test* 2009;9(4):333–357 <http://dx.doi.org/10.1080/15305050903352040>.
 58. Fisher WP. Rating scale instrument quality criteria. *Rasch Meas Trans Index* 2007;21(1):1095.
 59. Ataei S, Mahmud Z. Rasch-Andrich thresholds in engineering students' attitudes towards learning mathematics. *Advances in Mathematics and Statistical Sciences*;410–417. Retrieved from (<http://www.wseas.us/e-library/conferences/2015/Dubai/MCSS/MCSS-53.pdf>); 2015.
 60. Gibbons C. Stress, coping and burn-out in nursing students. *Int J Nurs Stud* 2010;47(10):1299–1309 <http://dx.doi.org/10.1016/j.jnurstu.2010.02.015>.
 61. Guthrie Black D, Bagalkote, H, et al. Psychological stress and burnout in medical students: a five-year prospective longitudinal study. *J R Soc Med* 1998;91(5):237–243 <http://dx.doi.org/10.1177/014107689809100502>.

62. Popa-Velea O, Diaconescu L, Mihăilescu, A, et al. Burnout and its relationships with alexithymia, stress, and social support among Romanian medical students: a cross-sectional study. *Int J Environ Res Public Health* 2017;14(6):560 <http://dx.doi.org/10.3390/ijerph14060560>.
63. Campos JADB, Maroco J. Maslach Burnout Inventory-Student Survey: portugal-brazil cross-cultural adaptation. *Rev Saude Publ* 2012;46(5):816–824 <http://dx.doi.org/10.1590/S0034-89102012000500008>.
64. Campos JADB, Jordani PC, Zucoloto, ML, et al. Burnout syndrome among dental students. *Rev Bras Epidemiol* 2012;15(1): 155–165 <http://dx.doi.org/10.1590/S1415-790X2012000100014>.
65. Hu Q, Schaufeli WB. The factorial validity of the Maslach Burnout Inventory-Student Survey in China. *Psychol Rep* 2009;105(2): 394–408 <http://dx.doi.org/10.2466/PRO.105.2.394-408>.
66. Maroco J, Campos JADB. Defining the student burnout construct: a structural analysis from three burnout inventories. *Psychol Rep* 2012;111(3):814–830 <http://dx.doi.org/10.2466/14.10.20.PRO.111.6.814-830>.
67. Costa EFDO, Santos SA, Santos, ATRDA, et al. Burnout syndrome and associated factors among medical students: a cross-sectional study. *Clinics* 2012;67(6): 573–580 [http://dx.doi.org/10.6061/clinics/2012\(06\)05](http://dx.doi.org/10.6061/clinics/2012(06)05).