JOURNAL OF WELLNESS

ORIGINAL RESEARCH

# Factor Structure and Measurement Invariance of the Maslach Burnout Inventory in Emergency Medicine Residents

Tim P. Moran, PhD[1], Nicole Battaglioli, MD[1], Simiao Li-Sauerwine, MD, MS[2]

## Abstract

**Introduction:** Emergency medicine residents suffer from high rates of occupational burnout. Recent research has focused on identifying risk and protective factors for burnout as well as targets for intervention. This research has primarily employed the Maslach Burnout Inventory to evaluate burnout in this population. Factor analytic work has identified three underlying factors measured by the Maslach Burnout Inventory: Emotional Exhaustion, Depersonalization, and Personal Accomplishment. However, this three-factor structure has not been evaluated in emergency medicine residents. Furthermore, its structural equivalence has not been demonstrated across commonly-studied risk factors, such as gender and year of post-graduate training. In the present study, we evaluated the structure of the Maslach Burnout Inventory in emergency medicine residents as well as its measurement invariance across gender and post-graduate year.

**Methods:** 1522 emergency medicine residents (21.1% of all US residents from 78.1% of US residency programs) were recruited as part of the 2017 National EM Resident Wellness Survey and completed the Maslach Burnout Inventory – Human Services Survey. The factor structure and measurement invariance across both respondent gender and post-graduate year were evaluated using a series of confirmatory factor analyses. Exploratory analyses evaluated whether burnout scores differed across men/women and post-graduate years 1, 2, and 3+ using a structural equation model.

**Results:** The three-factor structure was observed after minor modifications which replicated in cross-validation. This structure was invariant across both gender and post-graduate year at the configural, metric, and scalar levels. Emotional exhaustion scores were higher for female residents and scores on all of the MBI scales indicated greater burnout for more advanced residents.

**Conclusion:** These results indicate the Maslach Burnout Inventory is fully structurally equivalent across gender and post-graduate year and further validates its use in this population. Secondary evaluations of the latent means revealed that female residents tend to have higher scores on Emotional Exhaustion and that scores on all factors tend to worsen as trainees progress through their residency.

DOI: 10.18297/jwellness/vol2/iss2/2
Received Date: May 11, 2020
Accepted Date: July 15, 2020
Publication Date: Aug 25, 2020

Website: https://ir.library.louisville.edu/jwellness/

Recommended Citation: Moran, Tim P.; Battaglioli, Nicole; Li-Sauerwine, Simiao (2020) " Factor Structure and Measurement Invariance of the Maslach Burnout Inventory in Emergency Medicine Residents" Journal of Wellness: Vol. 2 : Iss. 2, Article 2.

Affiliations: [1] Emory University Department of Emergency Medicine, [2]The Ohio State University Department of Emergency Medicine

## INTRODUCTION

### Occupational Burnout in the Medical Profession

In recent years, there has been increased concern regarding physician wellbeing given that multiple studies have identified that American physicians suffer from high rates of occupational burnout [1-4]. In addition to increasing the risk of turnover, psychiatric conditions, and suicidality for physicians, burnout also results in reduced efficiency and quality of care [5-10]. Due to the unique demands of the emergency department's clinical environment, emergency medicine (EM) physicians appear to be at even greater risk of burnout. Peckman and Shanafelt et al. both found that over half (59% and 70%, respectively) of EM physicians reported feeling burned out [3-4]. Shanafelt et al., in particular, found that EM physicians reported the highest rates of burnout among the medical sub-specialties they surveyed [3].

Occupational burnout does not begin when a trainee becomes a practicing physician. Studies have found high rates of burnout in medical students, fellows, pediatric residents, surgical residents, radiology residents, neurosurgery residents, and emergency medicine residents [1, 11-14]. Similar to their attending counterparts, EM residents tend to suffer from higher rates of burnout compared to residents in other sub-specialties [2]. Understandably, researchers have begun trying to understand the risk and protective factors that influence burnout. In particular, several studies have identified possible demographic (e.g. gender, age/postgraduate training year [PGY]), personality (e.g. neuroticism), and institutional (e.g. hostile vs supportive mentorship) factors that influence burnout [15-21]. Such research has the potential to identify the trainees most likely to suffer from burnout and to identify targets of burnout interventions.

### Validity and Measurement Invariance

Given the scope of the problem and the associated stakes, it is imperative that researchers interested in physician wellness use well-validated instruments. "Validity", broadly, refers to whether scores derived from questionnaires (or other tests) measure the underlying trait that they purport to measure (e.g. whether a depression questionnaire actually measures some aspect of depression) and whether inferences based on those scores are justified [22-23]. Factorial validity, specifically, refers to the internal structure of a questionnaire and whether that

*Correspondence To: Tim Moran
Email: tpmoran@emory.edu

structure matches existing theories [24]. For example, one of the most common measures of burnout used to study medical providers is the Maslach Burnout Inventory – Human Services Survey (MBI-HSS). Factor analytic work has evaluated the internal structure of the MBI and suggests that it is composed of three underlying factors: Emotional Exhaustion (i.e. emotional depletion related to one's occupation), Depersonalization (i.e. negative affectivity and/or cynicism towards those encountered at one's occupation), and Personal Accomplishment (i.e. a sense of accomplishment related to one's occupation) [25-26]. Related to factorial invariance is the issue of measurement invariance. "Measurement invariance" refers to whether the internal structure is equivalent across different populations (e.g. is the three-factor structure present for both male and female respondents).

Establishing factorial validity and measurement invariance are critical for continued research on physician wellbeing. If the factor structure is not invariant across populations, then researchers cannot know whether observed differences between groups on the MBI reflect true differences in the underlying traits or whether observed differences reflect measurement bias because the MBI scales measure different traits across groups. For example, a hypothetical depression questionnaire might include an item related to the frequency of crying as an indicator of depression. However, if, for cultural reasons, men are less likely to cry even if depressed, this item would only be an indicator of depression for women [27]. Any comparison between men and women using this questionnaire would be contaminated with items that are not indicators of the characteristic being studied for both groups. As a result, it would not be possible to know whether any observed differences between men and women were due to true differences in depression, or due to measurement bias. With respect to the MBI, this questionnaire includes a number of items related to how individuals respond to stress (e.g. I feel burned out from my work). Given documented gender differences in stress responses and coping styles it is possible that male and female residents do not respond similarly to items on the MBI [28-29]. Similarly, perhaps as residents go through their programs and gain experience, skills and coping styles, certain behaviors cease to be indicative of stress. As with the example using gender, this would indicate that the questionnaire is no longer measuring the same underlying trait for residents in all years of training. In short, establishing measurement invariance across groups in a given population is critical when the goal is to compare scores between those groups. Without establishing measurement invariance, interpreting differences between groups and making decisions based on those findings would not be justified. In such a case, the resulting questionnaire scores would not be measuring the same underlying trait in both groups (i.e. the comparison would not be "apples-to-apples").

There are three types of invariance that are relevant for the present study: configural invariance, metric invariance, and scalar invariance. Configural invariance refers to determining whether the number of latent factors (e.g. Emotional Exhaustion, Depersonalization, and Personal Accomplishment on the MBI) and their relationship with the questionnaire items (e.g. "I feel burned out from my work" is part of the Emotional Exhaustion scale and not the other scales) are the same across groups. Metric invariance refers to determining whether the factor loadings for each questionnaire item are equivalent across groups. If metric invariance is not met, then some items are not equally relevant to the underlying factor across groups. Finally, scalar invariance refers to whether the item intercepts are equivalent across groups. The relationship between an underlying trait and a questionnaire item measuring that trait can be thought of as a regression equation. For this equation, scalar invariance tests whether the intercepts are equivalent across groups and metric invariance tests whether the slopes are equivalent across groups. If configural, metric, and scalar invariance are met, then comparing scores on the underlying factor is statistically permissible. Some texts also recommend evaluating residual invariance [30]. However, since the residual is not part of the latent mean, residual invariance is not necessary for comparing latent means [31].

### The Present Study

While the three-factor structure has been found in a number of populations (e.g. nurses, software engineers), the relatively new focus on resident physician burnout means that there is a dearth of psychometric data for this population [32-33]. Similarly, it is unknown whether the three-factor structure is invariant across subgroups (e.g. men vs women) within emergency medicine residents. The 2017 National EM Resident Wellness Survey provides an opportunity to examine the factorial validity and measurement invariance of the MBI. This is a large national survey in which 1522 respondents (21.1% of all EM residents from 78.1% of EM residency programs) completed MBI. The present study had three primary goals. First, we aimed to replicate the three-factor structure of the MBI in a large, national sample of emergency medicine residents. Second, we aimed to evaluate the evidence for measurement invariance across male and female residents and across postgraduate training years. Third, assuming that measurement invariance was achieved, this study aimed to compare burnout across male and female residents and across postgraduate training years.

## METHODS

### Study Design/Participants

The 2017 National EM Wellness Survey was conducted by a volunteer initiative from March 20-31, 2017. Emergency medicine residents were recruited to participate in the survey using multiple online sources, including the Academic Life in Emergency Medicine website (https://www.aliem.com),the Council of Emergency Medicine Residency Directors organizational listserv, the Emergency Medicine Residents' Association organization listserv, and through social media. The complete survey was hosted online using REDCap, a secure application for building and managing online surveys and databases. Respondents were compensated with a $5 Starbucks gift card and coupon codes to meal delivery services. Additionally, programs with more than 90% survey completion rates were entered into a lottery for a free pizza party and access to live-streamed recordings from a national emergency medicine education conference. Participant status as a current US emergency medicine resident was verified by obtaining and cross-referencing resident rosters from program directors, residency coordinators, or chief residents of accredited emergency medicine residency programs. Submissions were reviewed against program rosters to ensure resident status and that each resident responded only once. Submissions made by unconfirmed participants, duplicate submissions, and residents from dual or triple residency programs, such as combined emergency medicine/internal medicine programs, were excluded. All data were deidentified before analysis. This study was reviewed and approved by the Institutional Review Board.

### Measures

The primary measure of interest was the Maslach Burnout Inventory - Human Services Survey. This is a 22-item questionnaire that has been psychometrically decomposed into three underlying variables: Emotional Exhaustion, Depersonalization, and Personal Accomplishment [25]. The primary goal of the present study is to evaluate how well this three-factor structure fits the data in a population of emergency medicine residents. Respondents also provided basic demographic information (e.g. age, gender, years of post-graduate training) and responses to other factors thought to be related to wellness.

### Statistical Analysis

Continuous variables were described using medians and interquartile ranges. Categorical variables were described using frequencies and percentages. Primary data analysis consisted of a series of confirmatory factor analyses. As a first step, we evaluated the correlated three-factor model commonly used with MBI: Emotional Exhaustion, Depersonalization, and Personal Accomplishment. All three factors were allowed to correlate freely and all correlations between residuals were set to 0. However, for reasons described in more detail in the Discussion, factor structures for complex personality and attitude questionnaire can be difficult to replicate precisely [34]. For this reason, in the event that the initial model failed to achieve adequate model fit, an a priori decision was made to allow residuals with the largest modification indices (> 20) to correlate. Because this is an inherently exploratory procedure, it is prone to false positives/overfitting. To address this, the sample was first randomly divided into two groups (N = 761 per group) and the model was modified using modification indices in one subsample. The resulting model was then cross-validated in the other subsample. As is standard practice [31,35-36], multiple fit indices were used to evaluate model fit: the root mean-square error of approximation (RMSEA) should not exceed 0.08, the standard mean-squared residual (SRMR) should not exceed 0.1, the normed fit index (NFI) should be greater than 0.9, and the Tucker-Lewis index (TLI) should be greater than 0.9.

Once the model was identified, measurement invariance was evaluated separately for gender and PGY. The first model, the baseline model, allows all of the parameters to be estimated freely (i.e. does not constrain them to be the same for men/women or PGY1/PGY2/PGY3+). This tests configural invariance. If configural variance was established, then stricter equality restrictions were applied to assess other forms of measurement invariance [30]. In the second model, metric invariance was evaluated by constraining the factor loadings between the individual questionnaire items and their latent factor to be equal across men/women and PGY1/PGY2/PGY3+. The third and strictest test of measurement invariance evaluated scalar invariance by constraining the latent intercepts to be equivalent across men/women and PGY1/PGY2/PGY3+. At each step, the model is reevaluated to determine if the additional restriction was detrimental to model fit relative to the previous model. In order to evaluate measurement invariance, two fit indices are presented. Cheung & Rensvold's change in the comparative fit index (ΔCFI) has been found to be a good indicator of measurement non-invariance when it exceeds 0.01 [37]. While this provides a dichotomous decision making tool for evaluating fit, it does not describe how much constraining model parameters impacts model fit. To evaluate the magnitude of the effect on fit, we present Newson's $w^2$ [38]. Like the square of a Pearson correlation, this can be interpreted as a proportion of variance.

If the factor structures were found to be invariant, a secondary analysis was conducted to determine whether there were gender-based or PGY-based differences in burnout scores. For these comparisons, gender and PGY were added to the resultant model (see Schumacker & Lomax for a gentle introduction to structural equation modeling [39]). PGY was dummy coded with PGY1 as the referent category. The relationships between gender, PGY, and burnout were evaluated using standardized β coefficients, 95% confidence intervals, and p-values. Standardized β coefficients can be interpreted similarly to a correlation coefficient. Analyses were conducted using maximum likelihood in AMOS (v. 25, Armonk, NY).

## RESULTS

The final sample consisted of 1,522 residents. The sample had a median age of 30 years (IQR: 28-32). 879 (57.8%) residents identified as male and 643 (42.2%) identified as female. 523 (34.4%) were in their first year of residency (PGY1), 437 (28.8%) were in their second year of residency (PGY2), and 562 (36.9%) were in their 3+ year of residency (PGY3+). For the MBI subscales, the median responses were as follows: Emotional Exhaustion median = 25 (IQR: 18-33, Cronbach's α = 0.92), Depersonalization median = 14 (IQR: 9-19, Cronbach's α = 0.81), and Personal Accomplishment median = 37 (IQR: 32-42, Cronbach's α = 0.85).

### Initial Factor Structure

The fit indices for the overall model are presented in **Table 1**. The SRMR was acceptable; however, the other indices were poor-to-marginal. Thus, as described in the methods, the sample was randomly divided into two subsamples. In one sample, correlations between residuals with large modification indices were unconstrained and allowed to be freely estimated. This resulted in only 10 correlations (out of a possible 209 correlations) being freely estimated. As shown in Table 1, model fit was substantially improved in this modified model. Importantly, this same model was then cross-validated in the second subsample. Model fit was very similar, and acceptable, in the cross-validation sample (Table 1). Thus, this factor structure was used in the tests of invariance. Factor loadings for the initial and the final models are presented in **Table 2 (see next page)**.

Table 1: Model Fit for the Initial and Modified Models

| Model | NFI | TLI | CFI | RMSEA | SRMR |
|---|---|---|---|---|---|
| Model 1: No Correlated Residuals | .84 | .83 | .85 | .09 | .07 |
| Model 2: Correlated Residuals – Training Sample | .91 | .91 | .93 | .07 | .07 |
| Model 2: Correlated Residuals – Test Sample | .91 | .91 | .92 | .07 | .06 |

**Note.** NFI: Normed Fit Index; TLI: Tucker-Lewis Index; CFI: Comparative Fit Index; RMSEA: Root Mean Squared Error of Approximation; SRMR: Standardized Root Mean Squared Residual.

Table 2: Factor Loadings for the Initial and Final/Modified Models

| | Initial Model | | | Final/Modified Model | | |
|---|---|---|---|---|---|---|
| **Item** | **EE** | **DP** | **PA** | **EE** | **DP** | **PA** |
| I feel like I'm at the end of my rope | 0.68* | – | – | 0.69* | – | – |
| Working with people directly puts too much stress on me | 0.60* | – | – | 0.58* | – | – |
| I feel I'm working too hard on my job | 0.70* | – | – | 0.69* | – | – |
| I feel frustrated by my job | 0.76* | – | – | 0.75* | – | – |
| I feel burned out from my work | 0.86* | – | – | 0.88* | – | – |
| Working with people all day is really a strain for me | 0.65* | – | – | 0.65* | – | – |
| I feel fatigued when I get up in the morning and have to face another day on the job | 0.80* | – | – | 0.78* | – | – |
| I feel used up at the end of the workday | 0.82* | – | – | 0.78* | – | – |
| I feel emotionally drained at work | 0.84* | – | – | 0.82* | – | – |
| I feel recipients blame me for some of their problems | – | 0.44* | – | – | 0.43* | – |
| I don't really care what happens to some recipients | – | 0.59* | – | – | 0.54* | – |
| I worry that this job is hardening me emotionally | – | 0.82* | – | – | 0.83* | – |
| I've become more callous toward people since I took this job | – | 0.88* | – | – | 0.89* | – |
| I feel I treat some recipients as if they were impersonal objects | – | 0.62* | – | – | 0.58* | – |
| In my work, I deal with emotional problems very calmly | – | – | 0.56* | – | – | 0.48* |
| I have accomplished many worthwhile things in this job | – | – | 0.74* | – | – | 0.79* |
| I feel exhilarated after working closely with my recipients | – | – | 0.71* | – | – | 0.75* |
| I can easily create a relaxed atmosphere with my recipients | – | – | 0.66* | – | – | 0.60* |
| I feel very energetic | – | – | 0.55* | – | – | 0.56* |
| I feel I'm positively influencing other people's lives through my work | – | – | 0.72* | – | – | 0.72* |
| I deal very effectively with the problems of my recipients | – | – | 0.59* | – | – | 0.51* |
| I can easily understand how my recipients feel about things | – | – | 0.56* | – | – | 0.51* |

**Note.** EE – Emotional Exhaustion; DP – Depersonalization; PA – Personal Accomplishment
*p<.001

### Invariance by Gender

The first test of invariance was configural invariance for gender. This test allows all parameters to be estimated freely but constrains the number and configuration of factors to be the same across genders. As with the overall model, this resulted in acceptable model fit (NFI = 0.90, TLI = 0.91, RMSEA = 0.05, SRMR = 0.06), suggesting that a three-factor model with 10 correlated errors fits the covariance between these items approximately equally well for both men and women. Next, metric invariance was tested by constraining the factor loadings to be equal between men and women. The reduction in fit was negligible. ΔCFI was only .002 and $w^2$ was only .0002. Finally, we tested for scalar invariance by constraining the intercepts to be the same across men and women. This also resulted in a relatively negligible change in fit (ΔCFI = 0.008, $w^2$ = .005). That scalar invariance has been achieved allows one to compare the latent means across men and women.

### Invariance by PGY

Next we tested configural invariance for PGY. This test allows all parameters to be estimated freely but constrains the number and configuration of factors to be the same across training years. As with the overall model, this resulted in acceptable model fit: NFI = 0.90, TLI = 0.91, RMSEA = 0.04, SRMR = 0.06 suggesting that a three-factor model with 10 correlated errors fits the covariance between these items approximately equally well for both PGYs 1, 2, and 3+. Next, metric invariance was tested by constraining the factor loadings to be equal between training years. As with gender, the reduction in fit was negligible (ΔCFI = 0.001, $w^2$ = .0009). Finally, we tested for scalar invariance by constraining the intercepts to be the same across year of post-graduate training. This also resulted in a relatively negligible change in fit (ΔCFI = 0.001, $w^2$ = .002). That scalar invariance has been achieved allows one to compare the latent means across PGYs 1, 2, and 3+.

### Exploratory Comparison of Burnout Scores

Comparisons of the latent means on the burnout factors are presented in Table 3. Emotional Exhaustion was the only factor to differ significantly by gender - male residents had a lower latent mean than female residents. All burnout factors differed across training years. PGY2 and PGY3+ scored higher on Emotional Exhaustion and Depersonalization than PGY1. Similarly, PGY2 and PGY3+ scored lower on personal accomplishment than PGY1.

Table 3: Standardized Scale Differences Between Men/Women and PGY1/PGY2/PGY3+

| Characteristic | Scale | Standardized β | 95% CI | p |
|---|---|---|---|---|
| Gender | EE | -0.10 | -0.15, -0.05 | < 0.01 |
| | DP | 0.01 | -0.04, 0.06 | 0.72 |
| | PA | 0.03 | -0.03, 0.09 | 0.36 |
| PGY2 vs PGY1 | EE | 0.11 | 0.04, 0.17 | < 0.01 |
| | DP | 0.16 | 0.09, 0.22 | < 0.01 |
| | PA | -0.08 | -0.13, -0.02 | 0.02 |
| PGY3+ vs PGY1 | EE | 0.10 | 0.04, 0.16 | < 0.01 |
| | DP | 0.17 | 0.11, 0.23 | < 0.01 |
| | PA | -0.09 | -0.15, -0.02 | < 0.01 |

**Note.** PGY – Post-Graduate Year; EE – Emotional Exhaustion; DP – Depersonalization; PA – Personal Accomplishment; Standardized β – the standardized regression coefficient relating gender and PGY to the MBI scales; 95% CI – 95% Confidence Interval for the standardized β; p – p value for the standardized β

## DISCUSSION

In the present study, we examined the factor structure of the Maslach Burnout Inventory-HSS in US EM residents as well as its invariance across gender and years of post-graduate training. The initial CFA found that the simple three-factor structure resulted in inadequate model fit. However, when the residual terms with the most extreme modification indices were allowed to correlate in a random subsample of the data, this substantially improved fit. Importantly, when this same modified model was cross-validated using the other subsample (i.e. the half of the sample on which it was not developed), fit remained adequate. Such a finding is understandable given questionnaire methodology. It is tempting to assume that a questionnaire item actually evaluates only the underlying trait that it putatively evaluates (e.g. that an item measures emotional exhaustion and no other trait). However, it is difficult, if not impossible, to write perfect questionnaire items [34]. As a result, responses to each item may be influenced by any number of underlying traits, in addition to what it is meant to measure. As an example, the residual terms with the largest modification index belonged to the items "Working with people directly puts too much stress on me" and "Working with people all day is really a strain for me". In addition to measuring some aspect of emotional exhaustion with one's work, these items also measure how individuals respond to social interactions and "working with people". This additional shared content results in a correlation between responses to these items that exceeds what can be accounted for by the fact that they both measure burnout. Given the facts that 1) at least some of the misfit can be attributed to unavoidable artifacts in the item writing process, and 2) that the modified structure replicated in the test sample, the three-factor model seems to replicate in US EM residents.

This study also evaluated the invariance of the MBI-HSS across gender and years of post-graduate training. For both gender and PGY, measurement invariance was found across all three levels of invariance testing: Configural invariance demonstrated that men/women and PGY1/PGY2/PGY3+ were characterized by the same number and configuration of factors. Metric invariance demonstrated that the relationship between the underlying factors and the questionnaire items were not different across men/women and PGY1/PGY2/PGY3+. Finally, scalar invariance demonstrated that the intercepts did not differ between men/women and PGY1/PGY2/PGY3+. This finding is critical because it indicates that burnout, as measured by the MBI, can be conceptualized similarly across men/women and PGY1/PGY2/PGY3+. From a practical perspective, this finding also substantiates comparisons between men/women and PGY1/PGY2/PGY3+ – in essence, these findings suggest that such a comparison would be "apples to apples".

Given that measurement invariance held in this sample, we conducted a secondary analysis which examined whether burnout differed by gender or by PGY. This analysis found that women reported more Emotional Exhaustion than their male counterparts and that men and women reported similar levels of Depersonalization and Personal Accomplishment. With respect to training, burnout increased following PGY1. PGY2 and PGY3+ residents reported more Emotional Exhaustion and Depersonalization. Perhaps counter-intuitively, PGY2 and PGY3+ residents also reported less personal

accomplishment. Any interpretation of this finding would be somewhat speculative. It may be related to the Dunning-Kruger effect [40]. That is, individuals with more skill/experience in a given domain are generally more aware of their limitations than those with little skill/experience in that domain.

## LIMITATIONS

In the present study, we allowed residual terms to correlate in order to improve model fit. While this resulting factor structure replicated in another subsample, the changes made to the model were based on modification indices rather than a thorough understanding of the residual correlation structure. As noted earlier, shared content unrelated to the underlying trait of interest (e.g. social interactions vs. emotional exhaustion) can result in correlations between items higher than what would be explained by the trait of interest. However, this is not limited to shared content. Similar content, wording, reading levels and any number of other characteristics can result in residual correlations. Thus, we do not regard the final model as the only possible model to fit the data [34]. A more thorough understanding of how item wording can influence correlations in a given population will be needed to test "the true model". The survey, though national, did not record race and ethnicity from respondents. Thus, it is not clear how representative the sample was of US EM residents with respect to race and ethnicity. It was also not possible to test for measurement invariance across race/ethnicity. Similarly, this survey did not allow for a non-binary response to gender thereby making it impossible to evaluate measurement invariance in those who do not identify as male or female. Finally, how individuals respond to stress, gender difference in coping style, and how training influences stress/coping can be culturally bound. Thus, these findings may not replicate in residency programs outside of the United States.

## CONCLUSION

The present results suggest that the three-factor structure of the MBI-HSS is applicable to US EM residents. Importantly, this factor structure was invariant across male and female residents as well as residents in their PGY1, PGY2, and PGY3+ years of training. These findings are timely given both the increased recognition of the high rates of burnout in EM and the recent focus on ameliorating burnout in this population. As a secondary finding, these results suggest that female residents experience more emotional exhaustion than their male counterparts and that PGY2 and PGY3+ residents experience more burnout on all subscales than their PGY1 counterparts.

Acknowledgements: The authors acknowledge Michelle Lin, MD of ALiEM for her support and guidance, and Arlene Chung MD, Sarah Mott MD, Dan Robinson MD, and Matthew Melamed, MD MPH for their roles in the initial 2017 National Burnout Survey.

Funding: The authors acknowledge US Acute Care Solutions, which sponsors the Wellness Think Tank, and ALiEM for purchasing licensing rights to the Maslach Burnout Inventory tool; and Hippo Education, Emergent Medical Associates, and ALiEM for funding incentives for survey completion.

Conflicts of Interest: The authors disclose that SLS is the Chief Academic Officer and NB is the former Chief Operating Officer of the ALiEM Wellness Think Tank. The authors disclose no conflicts of interest, financial or otherwise, which could be perceived as potential sources of bias.

## REFERENCES

1. Elmore LC, Jeffe DB, Jin L, Awad MM, Turnbull IR. National Survey of Burnout among US General Surgery Residents. J Am Coll Surg. 2016 Sep;223(3):440–51.
2. Lin M, Battaglioli N, Melamed M, Mott SE, Chung AS, Robinson DW. High Prevalence of Burnout Among US Emergency Medicine Residents: Results From the 2017 National Emergency Medicine Wellness Survey. Ann Emerg Med. 2019 Nov;74(5):682–90.
3. Peckman C. (2017). Race and ethnicity, bias and burnout. Medscape Lifestyle Report 2017. Retrieved from https://www.medscape.com/features/slideshow/lifestyle/2017/overview
4. Shanafelt TD, Hasan O, Dyrbye LN, Sinsky C, Satele D, Sloan J, et al. Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014. Mayo Clin Proc. 2015 Dec;90(12):1600–13.
5. Kuhn G, Goldberg R, Compton S. Tolerance for uncertainty, burnout, and satisfaction with the career of emergency medicine. Ann Emerg Med. 2009 Jul;54(1):106–113.e6.
6. Shanafelt TD, Bradley KA, Wipf JE, Back AL. Burnout and self-reported patient care in an internal medicine residency program. Ann Intern Med. 2002 Mar;136(5):358–67.
7. Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. Lancet. 2009 Nov;374(9702):1714–21.
8. Lemkau JP, Purdy RR, Rafferty JP, Rudisill JR. Correlates of burnout among family practice residents. J Med Educ. 1988 Sep;63(9):682–91.
9. Oreskovich MR, Kaups KL, Balch CM, Hanks JB, Satele D, Sloan J, et al. Prevalence of alcohol use disorders among American surgeons. Arch Surg. 2012 Feb;147(2):168–74.
10. Martini S, Arfken CL, Churchill A, Balon R. Burnout comparison among residents in different medical specialties. Acad Psychiatry. 2004;28(3):240–2.
11. Attenello FJ, Buchanan IA, Wen T, Donoho DA, McCartney S, Cen SY, et al. Factors associated with burnout among US neurosurgery residents: a nationwide survey. J Neurosurg. 2018 Nov;129(5):1349–63.
12. Dyrbye LN, West CP, Satele D, Boone S, Tan L, Sloan J, et al. Burnout among U.S. medical students, residents, and early career physicians relative to the general U.S. population. Acad Med. 2014 Mar;89(3):443–51.
13. Guenette JP, Smith SE. Burnout: Prevalence and Associated Factors Among Radiology Residents in New England With Comparison Against United States Resident Physicians in Other Specialties. AJR Am J Roentgenol. 2017 Jul;209(1):136–41.
14. Kemper KJ, Wilson PM, Schwartz A, Mahan JD, Batra M, Staples BB, et al. Burnout in pediatric residents: comparing brief screening questions to the Maslach Burnout Inventory. Acad Pediatr. 2019 Apr;19(3):251–5.
15. Manderino MA, Brown M, Peters M, Wirtz R.

Sources of stress for nurse practitioners. J Am Acad Nurse Pract. 1994 Apr;6(4):155–9.
16. Bell RB, Davison M, Sefcik D. A first survey. Measuring burnout in emergency medicine physician assistants. JAAPA. 2002 Mar;15(3):40–2.
17. Browning L, Ryan CS, Thomas S, Greenberg M, Rolniak S. Nursing specialty and burnout. Psychol Health Med. 2007 Mar;12(2):248–54.
18. Ashooh MP, Barnette K, Moran TP, O' Shea J, Lall MD. Advanced practice provider burnout in a large urban medical center. Adv Emerg Nurs J. 2019 Jul/Sep;41(3):234–43.
19. Maslach C, Leiter MP. Early predictors of job burnout and engagement. J Appl Psychol. 2008 May;93(3):498–512.
20. Maslach C, Jackson SE. The Role of Sex and Family Variables in Burnout. Sex Roles. 1985;12(7-8):837–51.
21. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Dev Rev. 2016 Sep;41:71–90.
22. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull. 1955 Jul;52(4):281–302.
23. Kane MT. Current Concerns in Validity Theory. J Educ Meas. 2001;38(4):319–42.
24. Piedmont RL. Factorial Validity. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research. Dordrecht: Springer; 2014. https://doi.org/10.1007/978-94-007-0753-5_984.
25. Maslach C, Jackson SE. Maslach Burnout Inventory. Palo Alto (CA): Consulting Psychologists Press; 1981.
26. Leiter MP, Durup J. The discriminant validity of burnout and depression: a confirmatory factor analytic study. Anxiety Stress Coping. 1994;7(4):357–73.
27. Langelaan S, Bakker AB, van Doornen LJ, Schaufeli WB. Burnout and Work Engagement: Do Individual Differences Make a Difference? Pers Individ Dif. 2006;40(3):521–32.
28. Bangasser DA, Valentino RJ. Sex differences in stress-related psychiatric disorders: neurobiological perspectives. Front Neuroendocrinol. 2014 Aug;35(3):303–19.
29. Matud MP. Gender differences in stress and coping styles. Pers Individ Dif. 2004;37(7):1401–15.
30. Meredith W. Measurement Invariance, Factor Analysis, and Factorial Invariance. Psychometrika. 1993;58(4):525–43.
31. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational Research. 2000;3(1):4–70.
32. Kalliath TJ, O'Driscoll MP, Gillespie DF, Bluedorn AC. A test of the Maslach Burnout Inventory in three samples of healthcare professionals. Work Stress. 2010;14(1):35–50.
33. Taris TW, Schreur PJ, Schaufeli WB. Construct validity of the Maslach Burnout Inventory-General Survey: A two-sample examination of its factor structure and correlates. Work Stress. 1999;13(3):223–37.
34. Hopwood CJ, Donnellan MB. How should the internal structure of personality inventories be evaluated? Pers Soc Psychol Rev. 2010 Aug;14(3):332–46.
35. Kline RB. Principles and practice of structural equation modeling. New York: Guilford Press; 1998.
36. Schreiber JB, Nora A, Stage FK, Barlow EA, King J. Reporting structural equation modeling and confirmatory factor analysis results: A review. J Educ Res. 2006;99(6):323–38.
37. Cheung GW, Rensvold RB. Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. Struct Equ Modeling. 2002;9(2):233–55.
38. Newsom JT. Longitudinal Structural Equation Modeling: A Comprehensive Introduction. Routledge (NY); 2015. https://doi.org/10.4324/9781315871318.
39. Shumaker RE, Lomax RG. A Beginner's Guide to Structural Equation Modeling. NJ: Lawrence Erlbaum Associates, Inc.; 2004. https://doi.org/10.4324/9781410610904.
40. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol. 1999 Dec;77(6):1121–34.