The Development & Validation of the Neurozone® Burnout Index

Introduction

While there is **no consensus** on the measurement of occupational burnout, Maslach and Jackson's conceptualization of the syndrome remains the most widely accepted, with the **Maslach Burnout Inventory** (MBI) the most commonly used measure. According to this conceptualization, burnout consists of three dimensions: emotional exhaustion (weariness and fatigue related to the psychological efforts made at work), cynicism and depersonalization (detachment, indifference, and/or negative attitudes towards work and the people in it), and reduced personal achievement (a decrease in productivity and/or a poor self-evaluation of one's abilities to perform). The MBI is a generic measure of burnout, meaning it is not tailored to, nor does it differentiate between, one or more specific occupations. Several other valid, generic measures of burnout are used across the globe, such as the **Copenhagen Burnout Inventory** (CBI), the **Oldenburg Burnout Inventory** (OBI), the **Burnout Clinical Subtypes Questionnaire** (BCSQ), and the **Burnout Assessment Tool** (BAT). Some of these tools (e.g., the OBI) measure one or two dimensions captured by the MBI, while others (e.g. the CBI) measure alternative dimensions such as the various contexts in which burnout presents.

Common to all of these measures is their relative lengthiness. For example, the MBI consists of 16 items; the BCSQ consists of 36 items, and the BAT consists of 33 items. Additionally, the measures have all been developed in high-income settings, and while some (e.g. the MBI) have been adapted for use in other socioeconomic and cultural contexts (e.g., China, Arabia), further validation of burnout measures in low-to-middle-income countries like South Africa is warranted. Thus, a gap exists in the measurement of burnout. There is a need for a much briefer (though conceptually sound and psychometrically robust) instrument that can apply across diverse occupational, income, and cultural contexts.

Scale Development

Brief Burnout Measure

As mentioned above, there have been several well-validated measures of burnout developed over the past several decades. However, we identified a need for a brief, reliable, and valid measure of burnout that could be completed quickly and conveniently in diverse employment contexts and socio-economic settings. With these aims in mind, we conducted a thorough and up-to-date review of current empirical evidence describing the critical components that make up burnout. Upon conclusion of the review, a list of relevant constructs were identified and operationalised by an expert panel of psychologists, and finally populated with their respective items. Data collection on the Neurozone® Burnout Index commenced in 2021.

Pilot Study

When the initial sample size surpassed 200 after commencing data collection, we decided to run a pilot factor analysis in order to determine whether any potential changes/additions need to be made to the scale before continuing data collection. Results at the time, with a sample size of 212, revealed that a one-factor solution was the most appropriate fit for the data. All inter-item correlations were significant and sufficiently large, with no problematic multicollinearity.

Furthermore, the Kaiser-Meyer-Olkin (KMO) statistic as a measure of sampling adequacy confirmed that the sample size was adequate, while Bartlett's test of sphericity confirmed that the extraction method used was suitable for the data. Finally, the one-factor solution was found to be highly reliable with a Cronbach's alpha value of 0.844. Based on these results, we decided to continue data collection on the scale in its current form.

Current Study

The Sample

Data collection for the current study took place from 2021 to 2024. The total sample consisted of 1520 employed individuals across all major industries, departments, job levels, and highest levels of education. The average age of the sample was 42 years (range = 18–86). The gender distribution of the total sample consisted of 54% women and 45% men, while 1% of individuals identified as non-binary or elected not to disclose their gender identity.

Methods

Exploratory & Confirmatory Factor Analysis

Due to the large sample size (1520), the decision was made to randomly split the dataset in half and to perform exploratory factor analysis (EFA) on the first subset and confirmatory factor analysis (CFA) on the second subset. This is a robust approach in terms of determining and confirming a reliable and stable factor structure: By conducting EFA on one half of the dataset, one is able to explore the underlying factor structure without preconceived predictions, allowing the factors to emerge from the data mathematically. Subsequently, performing CFA on the other half of the dataset, by specifying the factor structure from EFA, enables one to statistically confirm whether the emergent factor structure replicates well in an independent dataset (i.e. the second, CFA subset). This approach also ensures homogeneity in terms of sample characteristics across the two datasets, since they were collected from the same population; however, the data points are, importantly, independent of one another and can therefore be used in separate, comparative analysis.

Data Preprocessing & Analysis

All data management and analyses were done using R Statistical Computing Software Version 2023.09.0+463. Due to the ordinal nature of the data (scores are derived from Likert scale responses), we computed a polychoric (as opposed to Pearson) correlation matrix. This was coupled with using the ordinary least squares (OLS) estimator with the 'minres' method for extraction to further accommodate the ordinal nature of the data. For CFA, we used diagonally weighted least squares (DWLS) as an estimator. The DWLS is a robust estimator typically used in CFA in the context of ordinal data and polychoric correlation matrices.

Results

Exploratory Factor Analysis

Examining the scree plot along with its associated eigenvalues revealed that a one-factor solution appears to be the most appropriate fit for the data. We also examined the factor loading for each item, which was very large in each instance (range = 0.662–0.865). Based on these results, we retained all items (generally, a retention cut-off of 0.500 is considered very robust). The cumulative variance explained was 60%. The KMO statistic as a measure of sampling adequacy was 0.820, which is considered very good. This result indicates that factor analysis was a suitable extraction method for this data.

One-Factor Solution

The one-factor solution of the Neurozone® Burnout Index (NBI) consisted of 6 items assessing various dimensions of burnout. These dimensions are inefficacy (feeling overwhelmed by work expectations and demands, accompanied by a experiencing a decline in work productivity), cynicism and detachment (tending to feel irritated around colleagues and/or clients, and to withdraw from people and activities in the workplace) and exhaustion (feeling worn out, fatigued, and emotionally exhausted).

Reliability Analysis

We evaluated the internal consistency of the one-factor solution by computing Cronbach's alpha. We did this to evaluate the extent to which the items making up the factor actually measure the same underlying concept (e.g., burnout). Higher Cronbach's α values (which typically range from 0 to 1) indicate greater internal consistency, suggesting that the items are more reliably measuring the same underlying construct. The ideal range for Cronbach's α values, however, is between 0.70 and 0.90.

Results show that Cronbach's α for the one-factor solution for the Neurozone® Burnout Index was 0.87, which is indicative of excellent reliability. Other metrics from reliability analysis support this: The corrected-item total correlations for all items were all large and well above the cut-off of 0.30. Furthermore, results also show that Cronbach's α did not increase in the event of any of the items being removed. This indicates that all the items included contribute to the reliability of the one-factor solution, while also eliminating possible concerns about scale redundancy.

Confirmatory Factor Analysis

We conducted confirmatory factor analysis on the one-factor solution derived from EFA. We computed multiple fit indices to assess both absolute and incremental goodness-of-fit. More specifically, we report on the chi-square statistic and its associated significance level, as well as the chi-square statistic divided by the degrees of freedom (X^2/df). The latter was included in order to circumvent the potential confounding effects of sample size and multivariate nonnormality on the chi-square results. Other absolute indices include the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Incremental fit indices include the comparative fit index (CFI) and the Tucker-Lewis Index (TLI). The a priori cut-offs for the various fit indices include $p < 0.05$ for the chi-square statistic, a value < 5 for the X^2/df statistic, < 0.06 for the RMSEA, < 0.08 for the SRMR, and > 0.95 for the CFI and TLI.

CFA Results

Initial examination of the results showed that the model is an overall good fit for the data; however, there was evidence of high covariance between certain items. Upon examination of the relevant items, we established that there is robust theoretical justification to specify the covariances in the model in two instances. Final results showed that all items loaded significantly and strongly on the factor with a range of 0.62 to 0.74. We calculated the participant-to-parameter ratio (PPR) as a measure of the complexity of the model relative to the amount of data available. A PRR of at least 5 to 10 is generally regarded as ideal. Results show that our PPR is 54.29, which is considered very robust and is indicative of stable parameter estimates and good model generalizability.

We also calculated several fit indices in order to evaluate how well the model structure fits the data. Firstly, with regard to absolute fit indices, results show that the chi-square statistic was not statistically significant (X^2 = 12.392, p = 0.088), which indicates that the one-factor solution is a good fit for the data. This is supported by the X^2/df statistic, which yielded a value of 1.770 (cut off: < 5). Secondly, the RMSEA, as another absolute fit index, shows that the model is a close fit for the data (0.046, CI lower = 0.035, CI upper = 0.057, p = 0.876; cut-off < 0.06), while the SRMR provides additional evidence for the close fit of the one-factor solution (0.032; cut-off: < 0.080). In addition, the incremental fit statistics also indicate that the model is a good fit for the data (robust CFI = 0.996; robust TLI = 0.991). The Kaiser-Meyer-Olkin (KMO) statistic as a measure of sampling adequacy was 0.82, which is considered very good. In addition, and importantly, the KMO statistics from EFA and CFA were the same (0.820), which indicates that the shared variance among variables and the latent factor structure is similar in both data subsets. This also provides support for the generalizability of the factor structure across datasets.

Validity Testing

In addition to having good content validity based on the scale development methodology, we also set out to demonstrate other, objective forms of validity: concurrent validity and convergent validity.

Concurrent Validity

Concurrent validity is a form of criterion validity that measures how well a new test compares to a well-established test. In this case, to test for concurrent validity, we needed to demonstrate that there is a statistically significant relationship between the NBI and an existing burnout measure. To this end, we assessed whether there was a significant, positive correlation between the 6-item Neurozone® Burnout Index and the existing, well-validated 19-item **Copenhagen Burnout Index**. Results showed a significant, large correlation (r = 0.810) between these two measures. Therefore, concurrent validity was (convincingly) confirmed.

Convergent Validity

Convergent validity, which is a form of construct validity, refers to the extent to which two measures that theoretically should be related are, in fact, statistically related. For example, in the case of burnout, we would expect there to be a significant *inverse* (negative) relationship with resilience, and a significant *positive* relationship with other mental health-related outcomes (e.g., depression, anxiety, and sleep disturbance). Indeed, results showed a significant, negative and large correlation with psychological resilience (as measured by the **Neurozone® Resilience Index**; r = -0.555). Additional results showed very large, significant, positive correlations with the following mental health outcomes: depression (r = 0.750), anxiety (r = 0.695), and sleep disruption (r = 0.744). Therefore, convergent validity was also (convincingly) confirmed.

Summary & Conclusion

We identified a need for a brief, valid, and reliable measure of burnout that can be completed quickly and conveniently in diverse employment contexts. We achieved this by running exploratory factor analysis (EFA) on one subset of the data, and confirmatory factor analysis (CFA) on the other subset. Results from EFA revealed a one-factor solution, a factor structure which was upheld by results derived from CFA. All absolute and incremental fit indices confirmed a very good, close fit of the model to the data. In addition, we demonstrated very strong content, concurrent, and convergent validity. Taken together, these results provide convincing evidence for the reliability, stability, validity, and generalizability of the one-factor solution of burnout as measured by the brief 6-item Neurozone® Burnout Index.

Dr Mariza van Wyk
Neuropsychologist & Senior Scientist

Dr Etienne van der Walt
Neurologist & CEO

June 2024

neurozone®

www.neurozone.com