Original Papers

Polish Psychological Bulletin
2023, vol. 54(3) 207–216
DOI: 10.24425/ppb.2023.148053

Teresa Chirkowska-Smolak[1], Tomasz Górecki[1], Mateusz Klakus[1]
Weronika Metzger[1], Martyna Szargan[1]

[1] *Adama Mickiewicza University, Poznań, Poland*

CORRESPONDING AUTHOR - Teresa Chirkowska-Smolak, chirko@amu.edu.pl

The factorial validity of the Maslach Burnout Inventory: Student Survey (MBI-SS) in Poland

ABSTRACT: This paper explores the psychometric properties of the Polish adaptation of the MBI-SS questionnaire, which measures academic burnout in students. The factorial structure of the tool was examined, and its validity was evaluated using a sample from Polish state universities (N=935).

Findings suggest that the MBI-SS possesses strong psychometric parameters. Data supports the tool's 3-factor structure and internal consistency. However, there are reservations regarding one item (MBI_6). Consequently, the 15-item version of the MBI-SS is recommended as a suitable tool for assessing burnout in Polish students.

Keywords: *students, burnout, validation, Maslach Burnout Inventory*

INTRODUCTION

For many years, the phenomenon of professional burnout has been recognized as a syndrome whose causes lie in the emotionally demanding contact of employees with other people, e.g., with patients, students, or clients. Over time, it has been recognized that burnout not only affects people working in social professions but is also a universal problem (Maslach, Schaufeli, Leiter, 2001) and can even affect people who are not employees (Schaufeli, Martinez, Pinto, Salanova & Bakker, 2002). Professional burnout syndrome has been included in the latest WHO Classification of Diseases and Disorders, ICD-11, as a problem related to professional functioning, which highlights the growing recognition of burnout's significance and its potential impact on both individuals and the broader community.

The issue of burnout is an important research topic in psychology and has been intensively studied since it was first presented in the psychological literature in the mid-1970s. In 2009 Schaufeli, Leiter, and Maslach estimated that over 6000 books, chapters, dissertations, and journal articles had been published on Burnout (Schaufeli, Leiter, Maslach, 2009). The most popular concept of burnout is a multidimensional model developed by Christina Maslach, which has been widely supported by empirical research, and it is on this approach that the WHO bases its definition. Maslach initially defined burnout as a "psychological syndrome involving emotional exhaustion, depersonalization, and a diminished sense of personal accomplishment that occurred among various professionals who work with other people in challenging situations" (Maslach, 1993, p. 19) and emphasized that burnout involves a loss of concern for another person. Currently, a more universal concept of burnout is used, since it has been recognized that it is a phenomenon that can affect representatives of various professions, not necessarily social ones. To describe the three basic dimensions of burnout, Maslach and colleagues used modified descriptions of burnout symptoms — emotional exhaustion was replaced with exhaustion in general, which refers to both emotional functioning and the physical condition of an individual; depersonalization was replaced with the more general term "cynicism", which was understood as a distanced attitude toward work, while the sense of lack of personal achievements was narrowed down

to the sense of reduced professional efficacy (Maslach, Leiter, 2008).

Exhaustion is not simply felt; rather, according to Maslach and Leiter (2008), it leads to actions that distance the individual emotionally and cognitively from work, arguably a way of coping with work overload. Attempts to distance oneself from various aspects of work (cynicism, depersonalization) are a direct response to exhaustion. This attitude manifests itself in negative, cynical, and overly defensive distancing and loss of idealism. The third component of burnout reflects the employee's self-esteem. It refers to feelings of incompetence, lack of professional achievement, and low productivity.

With the increasing knowledge of the determinants and mechanisms of burnout, researchers have pointed out that people burn out not only in professional settings (Stawiarska, 2016). Students were also identified as a particularly vulnerable social group (Hu, Schaufeli, 2009). Even though formally speaking, the academic activity of students is not professional work, it can be said, from a psychological point of view, that studying is a type of work because it is an organized and preimposed, structured and goal-oriented activity that is subject to external evaluation. Schaufeli and his team (Schaufeli et al., 2002) proposed that academic burnout could be defined analogously to occupational burnout, as suggested by the authors: "burnout refers to feeling exhausted because of study demands, having a cynical and detached attitude toward study, and feeling incompetent as a student" (p. 465).

However, despite the domination of the Maslach model of burnout in the field, a debate among scholars on the nature of burnout continues. This debate revolves around two interrelated issues: the dimensionality of burnout and its scope (Schaufeli, Leiter, Maslach, 2009). There is no consensus on student burnout among researchers: some researchers assume a three-factor conception of burnout (Salgado & Au-Yong-Oliveira, 2021; Schaufeli et al., 2002), and others assume that burnout consists of only the two key factors of exhaustion and cynicism, excluding the third factor of reduced professional efficacy (Alarcon et al., 2011). As for the scope of burnout, some scholars consider burnout as a generic, context-free phenomenon that may occur outside work, not only work-related (Kristensen et al., 2005).

The concept of burnout may be applied to activities that are similar to work but are outside the occupational context, such as, for example, in studies. As Schaufeli and Taris note, such activities may be exhausting and may also allow people to withdraw from them (2005, p. 260). In the last decade, there has been a growing interest among psychologists in student academic burnout (Portoghese et al., 2018; Rostami et al., 2014; Salgado & Au-Yong-Oliveira, 2021; Schaufeli et al., 2002; Simancas-Pallares et al., 2017; Turhan et al., 2021). Many of these researchers have struggled to find (or develop) a good measurement tool. To assess burnout in students, various instruments have been used, such as the students' version of the two-factor Oldenburg Burnout Inventory (Campos et al., 2012) or the Copenhagen Burnout Inventory (Campos et al., 2013). However, they are not as widely used as that developed by Schaufeli et al. (2002) MBI-SS. This article presents a questionnaire that may prove useful in the analysis of academic burnout among Polish students. The questionnaire is a modification of the world's most popular test for measuring academic burnout, i.e., the Maslach Burnout Inventory (MBI-GSS).

METHOD

Research problems and hypotheses

The main objective of the study was to prepare a Polish adaptation of the tool to measure academic burnout in students, i.e., the MBI-SS, which is a special version of the world's most popular questionnaire to measure professional burnout. Although the results of the validation of this tool in different countries have already been published, they show that further work is needed to confirm its value, as single items of this questionnaire raise doubts among researchers and are eliminated from the final version in a given language (Schaufeli et al., 2002; Hu and Schaufeli, 2009; Portoghese, Leiter, Maslach, Galletta, Porru, D'Aloja, Finco, Campagna,.., 2018). Above all, there is no Polish version that could be used in research in the area of mental health and well-being of students. Therefore, the main research problem was to confirm the good validity of this tool in the Polish version and to determine whether the three-factor structure is also applicable in Poland.

The Maslach Burnout Inventory (MBI) questionnaire for measuring burnout by Maslach, Jackson, Leiter, Schaufeli, and Schwab was developed in 1981 (Maslach, Jackson, Leiter, 1996). There are several versions of this tool, designed to survey people from different professions: for medical personnel (MBI-PM), for social professions in general (MBI-HSS), for teachers (MBI-ES) and a general version (MBI-GS) designed to study this phenomenon among representatives of all professions. They differ not only in the indication of the context to which they refer but also in their length (e.g., the PM version has 22 items, while the GS version has 16). All versions of the MBI questionnaire are licensed by MindGarden.

However, the publisher does not have the Polish adaptation of the MBI-SS student burnout measurement tool. The latest addition to the MBI group was developed by Schaufeli, Martinez, Pinto, Salanova and Bakker (2002). It is a modified version of the general questionnaire for all occupations (MBI-GS), where for example, the term "work" has been replaced by "study". The questionnaire is used in different countries in a 16-item version (as in the original) or a 15-item version, omitting one test item from the cynicism scale. This is the MBI_13 item, which was also questionable in the Polish adaptation of the MBI-GS developed by Chirkowska-Smolak and Kleka (2011).

General information on the student version is presented in the current test manual (the 4th edition of

the Manual) of 2017, but the authors emphasize that its psychometric properties has not yet been sufficiently documented. However, since the development of the scale and the first validation studies in three countries, Spain, Portugal, and the Netherlands (Schaufeli et al., 2002), increasingly more data have become available each year, along with successive language versions, providing researchers with the opportunity to make international comparisons.

Tool description

The scale, similar to the general version, which was the basis for the construction of the student version, consists of 16 items that are assigned to three subscales: exhaustion, associated with the demands of studying (EX, 5 items, e.g., "I feel emotionally drained by my studies cynicism, showing cynical attitudes and lack of identification with one's studies (CY, 5 items, e.g., "I doubt the significance of my studies") and professional/academic efficacy, i.e., the feeling that one is not competent in the role of student (EF, 6 items reverse scored, e.g., "In my opinion, I am a good student").

Respondents provided answers on a 7-point Likert scale ranging from 0 (never) to 6 (every day). High scores on the Exhaustion and Cynicism scales suggest the presence of burnout in the students surveyed, while in the case of Academic Efficacy, on which all items are positively worded, burnout is indicated by a low score. It should also be added that the authors make it clear that an overall score for burnout should not be calculated, e.g., as an average of all test items/items. The results of each subscale are analyzed separately — they create individual profiles for each respondent.

One of the important research goals related to burnout and its measurement tools is identifying the factor structure. Previous studies have explored the dimensionality of the MBI. Still, they have yielded inconsistent and conflicting results because some researchers identified various factor solutions (Faraci, 2018). Especially there are some doubts as far as the inclusion of professional efficacy is concerned. Some researchers have shown that lack of personal accomplishment works differently from the two other MBI dimensions, suggesting that lack of personal accomplishment might not be part of the burnout syndrome (Schaufeli and Taris, 2005). The analyses conducted by Pasikowski (1996) in the Polish language version of the MBI questionnaire for social professions and later by Chirkowska-Smolak and Kleka (2011) in relation to the general scale indicated a satisfactory value of the Polish versions of the tools and confirmed the three-factor structure and reliability of the tool; however, there were doubts about the relationship between the individual subscales. The main point is that between the first two subscales (exhaustion and cynicism), there are quite strong relationships, while they are weaker in the case of a lack of professional efficacy, yet these are the dimensions of the same phenomenon. Exhaustion and cynicism occur together, and reduced professional efficacy is observed much less frequently (Lee & Ashforth, 1996; Brennikmeijer, Van Yperen, 2003). This dimension was, in fact, later included in the model (Maslach, 1993). Some authors suggest that professional efficacy is related to individual characteristics (Bandura's sense of self-efficacy) rather than being a component of burnout (Shirom, 2003). The results of a study conducted by a team of Spanish researchers allow us to suggest that it is the lack of professional efficacy that may contribute to the development of burnout (Salanova, Peiró, & Schaufeli, 2002) and that, in this case, it should be treated as an independent variable rather than a component of burnout (Bresó, 2008). One may also wonder the extent to which the wording of the test items contributes to this or whether the reverse wording of the items for the third subscale does not account for the weaker correlation results between the subscales (Chirkowska-Smolak, 2012).

Research procedure

Work on the Polish version of the questionnaire proceeded in two stages: translation and validation of the tool.

Development of the Polish language version.

The MBI-SS Burnout Questionnaire is a licensed tool published by Mind Garden Inc. Before starting work on the Polish version, consent was obtained from the publisher, who accepted the conditions of the survey and allowed us to administer the MBI as an online survey via non-Mind Garden survey system.

First, translation was performed according to the procedure for translation of research tools from the area of quality of life and health, i.e., translation and retranslation performed by independent translators and comparison of the original version with the retranslation (WHO, 2016). However, this procedure was extended, as we also took into account the suggestions of Beaton et al. (2000) regarding the cultural adaptation of tests.

First, prior to translation, we made a preliminary assessment of the individual test items, whether they were appropriate for measuring burnout among students and whether there were any cultural differences, such as those related to differences in the education system, which needed to be taken into account.

Then, the original version (in English) was translated into Polish by the authors of the adaptation (translations were conducted independently). After a discussion about discrepancies and ambiguities between the different language versions, the Polish version of the questionnaire was established through consensus. The next step was to perform retranslation with the help of two translators who did not know the original version (blind translation) and translated the Polish version into English independently. Finally, the original version was compared with the two retranslated versions, and the differences were discussed with a psychologist working in the US. We wanted to ensure that the Polish-language version and the original version would be equivalent, taking into account any differences between higher education systems. Once we

were satisfied that the original version and the retranslations were equivalent, we moved to the next stage related to the development of the Polish language version, which was the initial evaluation by the target group.

For this purpose, we used the help of 25 psychology students specializing in work and organizational psychology who formed an expert panel. Each completed the questionnaire and evaluated the way in which individual test items were expressed, the comprehensibility of instructions and the way of answering on the proposed scale. Additionally, they could make comments regarding the Polish translation. In this way, we checked how participants might understand and interpret individual items. Then, it was possible to move on to the actual study to confirm the factorial structure of the questionnaire.

Organization of the validation study

To validate the tool, we checked its factorial structure, verified the fit of the three-factor model, and calculated the reliability of the tool (its internal consistency and test retest).

Students were informed about the ongoing study on well-being connected with their studies through social media (e.g., departmental profiles on FB). Informed consent was obtained from each respondent according to APA guidelines. Participation in the study was voluntary and anonymous, as no personal data were gathered. Respondents could withdraw from the study at any time and close the form without saving the results. Respondents were informed that they could contact the researchers if they felt uncomfortable or had questions or concerns and necessary contact details were provided. Additionally, participants were informed of the possibility of seeing the results by contacting researchers.

In the first step, only the MBI questionnaire was distributed to students. In the appropriate form's section, they were able to indicate their willingness to participate in a further stage of the study (concerning construct validity). A similar strategy was adopted after it was sent; participants could declare their willingness to participate in the planned retest (described in Section 3.2.). Retest was conducted after 3 months.

Measure

In the retest the following questionnaires were used:

– Copenhagen Psychosocial Questionnaire COPSOQ II (Pejtersen, Kristensen, Borg, Bjorner, 2010; Polish version: Baka, 2019): burnout (4 items) and self-efficacy beliefs (6 items) subscales. The COPSOQ II is designed to comprehensively measure psychosocial working conditions, such as demands, organization, and content of work, social relations, and leadership or work values. Respondents are asked to indicate their answers for all statements on a 5-point scale ('1. Always, 5. Never' for the burnout subscale and '1. To a very large degree, 5. To a very low degree' for the self-efficacy beliefs subscale);

– Brief Job Satisfaction Measure II (judge, Bono, Locke, 2000; Polish version: Chirkowska-Smolak, Grobelny, 2016): in a modified version for students, the term 'work' has been replaced by 'study'. Respondents are asked to indicate for each of 5 items on a 7-point scale (1. Strongly disagree, 7. Strongly agree).

Participants

We determined the required sample size *a priori*. The sample size was computed using recommendations in literature (Kim, 2005) when conducting a confirmatory factor analysis (CFA). We took into account model complexity: number of items per factor (5, 5, and 6), relationships between factors (average factor correlation 0.4), and factor loading (average factor loading 0.6). For alpha, a value of 0.01 was selected as the target probability, and a value of 0.99 was selected for power. We expected that CFA would confirm a good fit of the model to the observed data (expected CFI 0.95). The minimal sample size was calculated with a web calculator retrieved from http://wnarifin.github.io (Arifin, 2023) as 574.

The Polish tertiary education system is based on a three-tier structure of studies. These are first degree studies (usually 3-year studies, bachelor's degree studies), second degree studies (master's degree studies, 2-year studies) or comprehensive master's degree studies, which concern strictly defined majors, such as psychology, law, or medicine (5-6 years). Studies can be conducted either as full-time studies or extracurricular studies.

A total of 935 students from state universities in Poland participated in the actual study[1]. A total of 67% of the sample was female, and 28,6% was male. Most of the respondents were studying full-time (70,9%). Approximately half of the respondents (53%) were students of social sciences (e.g., psychology, pedagogy, sociology), and the others studied science and natural sciences (e.g., mathematics, chemistry, biology, geography, 23%), humanities (e.g., philology, cultural studies, 12%) and law and economics (e.g., law in business, management, 12%).

They were students of all years (1st–5th year), 53% were students at the bachelor's degree level, 16.1% at the master's degree level, and the rest were students at the comprehensive master's degree level (majoring in psychology or law).

In the retest participated sample of 129 students (48.4% were women). The average age of the participants was 21, most of them (62.1%) did not engage in additional student activities, and 69.4% were not employed. 52% of them were social science students, 33% - studied science and natural sciences.

RESULTS

Description of statistical variables

The mean values for EX, CY, and EF were at a moderate level, with students reporting their efficacy at the highest level and cynicism at the lowest level. The

[1] We share our dataset on OSF: https://osf.io/kuc7g/?view_only=250757b377cc42afbf605cd4fcccfb9c

Table 1. Descriptive statistics and reliability coefficients (Cronbach's alphas if item deleted) for the three MBI-SS subscales

Subscale and items	M	SD	Reliability
1. Exhaustion	3.28	1.77	0.90
EX_1	2.62	1.40	0.88
EX_2	2.95	1.46	0.87
EX_3	2.83	1.71	0.87
EX_4	2.79	1.64	0.87
EX_5	2.08	1.83	0.88
2. Cynicism	2.73	1.96	0.89
CY_1	1.83	1.91	0.85
CY_2	2.17	2.01	0.84
CY_3	3.45	1.84	0.92
CY_4	2.64	1.97	0.87
CY_5	2.22	2.01	0.86
3. Efficacy	3.47	1.40	0.83
EF_1	3.60	1.40	0.83
EF_2	2.37	1.64	0.80
EF_3	3.31	1.67	0.82
EF_4	3.46	1.60	0.81
EF_5	1.87	1.63	0.81
EF_6	3.04	1.55	0.80

coefficients for variability (54% for exhaustion, 72% for cynicism, and 48% for efficacy) indicate strong variability in the results, with respondents being the most diverse in terms of their cynical approach to their studies.

Reliability

The reliability of the tool was measured using an internal consistency test. Cronbach's alphas for EX, CY, and EF were 0.90, 0.89, and 0.83, respectively, indicating a high level of internal consistency for all three MBI-SS subscales.

Figure 1 shows the correlations between the variables. We can see that the variables are highly correlated within subscales. In addition, some correlation is observable between the variables forming the EX and CY subscales, especially between EX_5 and the variables that make up the CY subscale.

Another way to test the reliability of the adapted tool was to conduct a repeat measurement 3 months apart. In the end, 129 responses were obtained for the comparison of absolute stability. Table 2 shows the obtained Pearson's r correlation coefficients between individual subscales and for the entire questionnaire. The coefficients obtained range from $r = 0.56$ for the exhaustion subscale to $r = 0.81$ for personal efficacy.

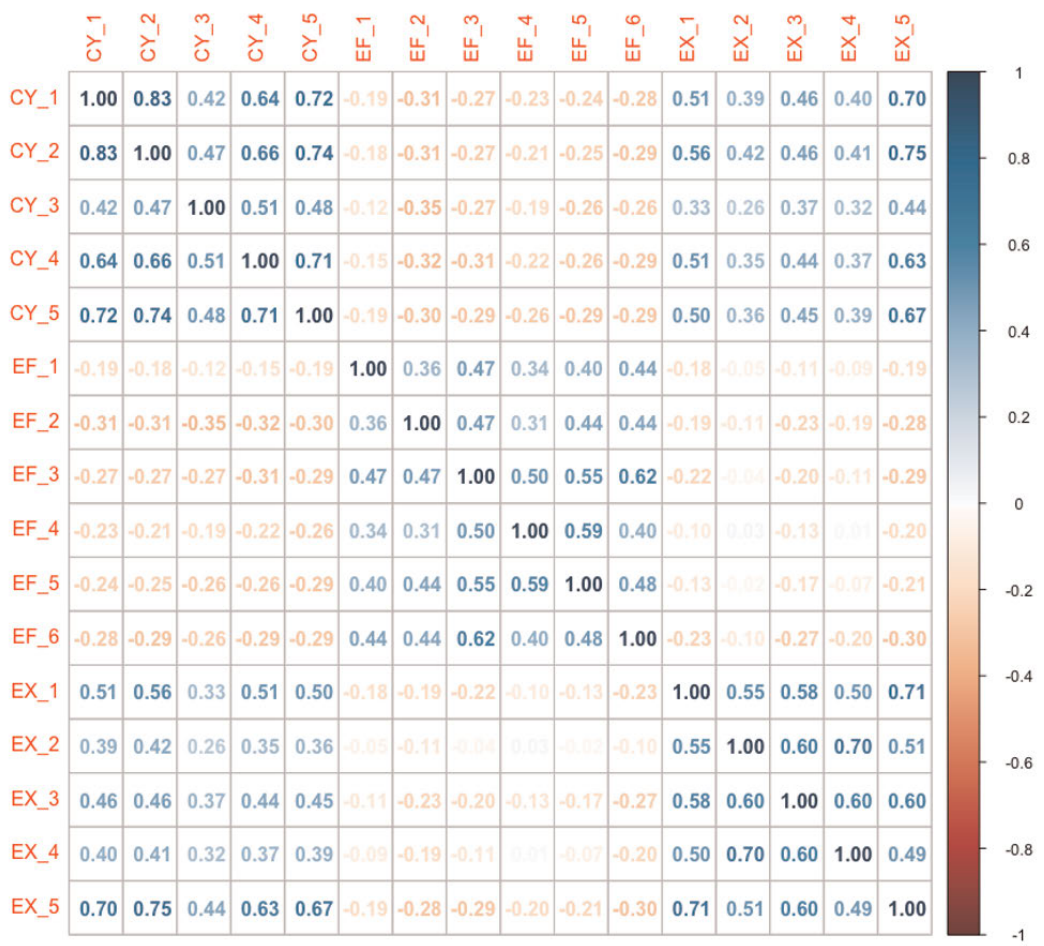

Figure 1. Correlation matrix

Table 2. Test-retest estimated by Pearson's r coefficient

Subscale	M_1	M_2	r	95% CIs
Exhaustion	3.07	3.25	0.56**	0.46, 0.69
Cynicism	2.56	2.81	0.74**	0.65, 0.81
Personal efficacy	4.09	3.27	0.81**	0.74, 0.86
MBI General Score	3.25	3.12	0.70**	0.60, 0.78

Note: $^{**}p < 0.01$

Confirmatory factor analysis

To confirm the three-factor structure of the scale that was corroborated for the original version, we performed confirmatory factor analysis using R project's library lavaan (Rosseel, 2012). Its results were also intended to allow us to decide on a test item that raised our earlier concerns.

To examine factorial validity, we assessed the fit of different models to identify which models best fit the data in our study. The goodness of fit was assessed using the χ^2 test statistic, the comparative fit index (CFI), and the Tucker–Lewis index (TLI). χ^2 tests the null hypothesis that the predicted model and observed data are equal. Because we want our predictions to match the actual data as closely as possible, we do not want to reject it (a nonsignificant result indicates a good model fit). Based on the literature (Hu & Bentler, 1999), the comparative fit index (CFI) and Tucker–Lewis index (TLI) cutoff scores should be above 0.95 for well-fitted models. Furthermore, a root mean square error of approximation (RMSEA) less than 0.06 and a standardized root mean square residual (SRMR) less than 0.08 indicate an acceptable model fit (Hu & Bentler, 1999). The chi-square values for all models were statistically significant ($p<0.01$) due to the large sample size (Tanaka, 1987; Babyak & Green, 2010).

By default, lavaan library (and Mplus) will always set the first loading of each factor to 1 (marker method). We checked multivariate normality. Mardia's test revealed multivariate non-normality. Base on this result our analyses were performed with maximum likelihood estimation with robust standard errors and a Satorra-Bentler scaled test statistic.

First, we assessed the fit of the one-factor (M1) solution, which assumes that all three aspects of burnout load are on one underlying dimension, as well as the fit of two- (M2) and three-factor solutions (M3), which assumes that two or three aspects of burnout are independent yet correlated factors. The two factor model (M2) result from combining two subscales: EX and CY, forming first factor and the PE subscale becoming second factor. Our results (Table 3) show that the fit of the three-factor solution (M3) appears to be somewhat better than that of the one- and two-factor solutions.

In the next step, we decided to test the model by leaving out the variable EX_5, which was highly correlated with CY scale (M4). This model obtained better results in terms of quality assessment. In this situation, we decided to perform additional tuning of the model. We used a method called modification indices (Sörbom, 1989). The modification index is an χ^2 value, by which model fit would improve if a particular path was added or a constraint freed. Values larger than 3.84 mean that the model could be 'improved', the p-value for the added parameter would be less than 5%, and values larger than 10.83 indicate that the parameter would have a p-value less than 0.1%.

Modification indices (MIs) are useful diagnostic tools in Confirmatory Factor Analysis (CFA) that help identify model misspecification and suggest potential improvements. They indicate how much the model's overall chi-square goodness-of-fit statistic would decrease if a particular parameter (e.g., a factor loading, covariance, or error covariance) were freely estimated instead of being fixed or constrained. In Table 4 we present MIs along with their corresponding Expected Parameter Change (EPC) values. MIs represent the expected decrease in the chi-square statistic if a parameter is freely estimated. Larger MIs suggest greater potential for model improvement. EPCs provide the expected change in the parameter value if it were freely estimated.

Table 3. Indices of overall fit for alternative factor structures

Model	χ^2(df)	CFI	TLI	RMSEA (90% CIs)	SRMR
M0: null model	8028.13 (120)	<0.001	<0.001	0.275 (0.270, 0280)	0.375
M1: one-factor model	2432.42 (104)	0.674	0.623	0.169 (0.163, 0.175)	0.137
M2: two-factor model	1235.25 (103)	0.839	0.812	0.119 (0.113, 0.125)	0.077
M3: three-factor model	821.50 (101)	0.897	0.878	0.096 (0.090, 0.102)	0.064
M4: modified three-factor model	509.50 (87)	0.930	0.916	0.079 (0.073, 0.086)	0.061
M5: final three-factor model	**141.15 (76)**	**0.989**	**0.985**	**0.033 (0.025, 0.042)**	**0.026**
M6: MGCFA configural model	215.58 (152)	0.989	0.985	0.033 (0.022, 0.043)	0.028
M7: MGCFA weak model	234.62 (168)	0.989	0.986	0.032 (0.022, 0.042)	0.033
M8: MGCFA strong model	317.46 (180)	0.977	0.973	0.044 (0.036, 0.052)	0.038

Table 4. Largest MIs for M4 (those MIs that have been added to the model are in bold).

Modification	Modification index (MI)	Expected Parameter Change (EPC)
CY_1~~CY_2	**132.91**	**0.62**
EX_2~~EX_4	**91.30**	**0.56**
EF_4~~EF_5	**84.54**	**0.52**
Cynicism=~EX_1	**81.07**	**0.34**
Efficacy=~EX_2	**48.36**	**0.37**
CY_4~~CY_5	45.91	0.39
Cynicism=~EX_2	38.69	-0.22
EX_1~~EX_4	35.53	-0.34
CY_3~~CY_4	33.85	0.40
CY_1~~CY_3	**31.26**	**-0.31**
CY_1~~CY_4	28.67	-0.28
Efficacy=~EX_3	24.43	-0.28
CY_2~~CY_4	24.10	-0.26
Efficacy=~EX_1	21.52	-0.27
EF_3~~EF_6	21.34	0.24
Cynicism=~EF_2	**21.18**	**-0.15**
EF_2~~CY_3	**20.80**	**-0.31**
CY_2~~CY_5	19.19	-0.24
EF_4~~EF_6	17.41	-0.21
Efficacy=~CY_3	17.14	-0.32
EX_4~~EF_4	**15.48**	**0.19**
Cynicism=~EX_4	14.39	-0.14
Exhaustion=~EF_2	13.97	-0.16
EF_2~~EF_4	13.46	-0.22
Exhaustion=~EF_6	**13.42**	**-0.13**
EX_1~~CY_2	**12.96**	**0.15**

We made modifications one at a time, starting with the largest MI. After each modification, re-ran the CFA, assessed the model fit, and checked the MIs again. We continued this iterative process until we achieved an acceptable model fit (all MIs less than 10.83).

Final M5 specification:

Exhaustion =~ EX_1 + EX_2 + EX_3 + EX_4 + EF_6
Efficacy =~ EF_1 + EF_2 + EF_3 + EF_4 + EF_5 + EF_6 + EX_2
Cynicism =~ CY_1 + CY_2 + CY_3 + CY_4 + CY_5 + EX_1 + EF_2
CY_1 ~~ CY_2
EX_2 ~~ EX_4
EF_4 ~~ EF_5
EF_2 ~~ CY_3
EX_1 ~~ CY_2
CY_1 ~~ CY_3
EX_4 ~~ EF_4

We checked also MIs for M6, M7 and M8. However, no additional changes have been suggested for these models.

Using different observed variables in different models makes the chi-square difference test results uninterpretable. This test requires that the models be nested, one a special case of the next. Models with different observed variables are not nested. Hence, in our situation is not possible to compare using this method models M3 (model with Ex5) and M4 (model without Ex5). Due to the frequent challenge of determining if non-nested models overlap, we generally employ a two-phase testing approach outlined by Vuong (1989). In the initial phase, we assess if the two models can be differentiated from each other, which is feasible if they overlap. In the subsequent phase, we evaluate if the fits of both models are equivalent. From the first test (the variance test) we obtained p-value 3.33E-9. Hence, M3 and M4 are distinguishable. Thus, we can move on to the second (non-nested likelihood ratio test) test. For alternative "M4 fits better than M3" we obtained p-value <0.001. Exactly the same conclusion is reached when comparing the M5 and M3 models. M5 fits better than M3.

Finally, we evaluated the obtained model (M5) using the goodness-of-fit indices. We obtained a CFI value of 0.983 and a TLI value of 0.976, which are good fit indices for internal validity. The value of RMSEA is 0.041, which indicates a good fit. Similarly, the SRMR value of 0.038 is below the cutoff score of 0.08 and indicates a good fit. The final model is shown in Figure 2.

To provide evidence of MBI-SS measurement invariance across different groups, we conducted multigroup CFA (MGCFA) across genders. Before we compare the values of latent means across multiple groups, we first need to establish measurement invariance. Testing for measurement invariance involves a fixed sequence of model comparison tests. A typical sequence involves three models:

1. Configural invariance (M6). The same factor structure is imposed on all groups.
2. Weak (metric) invariance (M7). The factor loadings are constrained to be equal across groups.
3. Strong (scalar) invariance (M8). The factor loadings and intercepts are constrained to be equal across groups.

We prepared the three models and compared them with each other. It turned out that the first of the three models (M6 in Table 7) is comparable to the next two models. Therefore, there is no reason to think that gender affects the results. In addition, we can note that this model is inferior model to the proposed M5 model (all quality measures are inferior), which did not account for gender.

We assumed that the factorial structure of the MBI-SS is invariant across the samples of females and males. However, compared to previous studies that used the MBI-SS, we expected no such differences; many researchers reported no significant effect of sex or that the factorial structure of the MBI-SS also was invariant across females

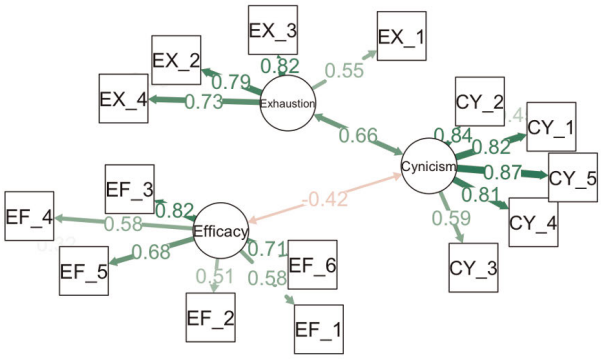

Figure 1. Confirmatory factor analysis plot

and males (Portoghese et al., 2018; Simancas-Pallares et al., 2017; Turhan et al., 2021).

The ANOVA tables below are for the models that can be so compared. The Tables 5-7 show that the M0-M3 models all differ significantly. We, therefore, choose the M3 model. Models M4 and M5 are also significantly different, so we choose model M5 from these. Models M6 and M7 do not differ significantly, but M7 and M8 do. We, therefore, choose M6 from this group of models.

Construct validity

The validity of the MBI-SS was measured in two ways. First, MBI scores were correlated with satisfaction from studies, the outcome that had been hypothesized to be related to burnout. Second, to establish convergent validity of the MBI, burnout scores were correlated with responses on another measure of burnout — the subscale of the COPSOQ II.

To determine convergent validity, the three dimensions of the adapted questionnaire were subjected to correlation analysis with the general level of burnout [GLB], self-efficacy beliefs [SEB], and satisfaction with studies [SwS] (Table 8).

The obtained results indicate a statistically significant correlation between the analyzed variables at the $p < 0.01$ level. Dimensions of exhaustion and cynicism correlate strongly and negatively with self-efficacy ($r = -0.29$ and $r = -0.28$) and negatively and significantly with satisfaction with studies ($r = -0.17$ and $r = -0.39$). The opposite is found for the Personal Efficacy subscale, which has positive, medium correlations ($r = 0.51$ and $r = 0.46$). Moreover, the higher the overall burnout score (GLB), the higher the score on the exhaustion and

Table 5. ANOVA table comparing the M0-M3 models.

Model	χ^2(df)	Difference in χ^2	p-value
M3: three-factor model	821.50 (101)		
M2: two-factor model	1235.25 (103)	413.75	0.00
M1: one-factor model	2432.42 (104)	1197.17	0.00
M0: null model	8028.13 (120)	5595.71	0.00

Table 6. ANOVA table comparing the M4-M5 models.

Model	χ^2(df)	Difference in χ^2	p-value
M5: final three-factor model	141.15 (76)		
M4: modified three-factor model	509.50 (87)	368.35	0.00

Table 7. ANOVA table comparing the M6-M8 models.

Model	χ^2(df)	Difference in χ^2	p-value
M6: MGCFA configural model	215.58 (152)		
M7: MGCFA weak model	234.62 (168)	19.04	0.29
M8: MGCFA strong model	317.46 (180)	82.84	0.001

Table 8. Pearson correlation coefficients for validity of MBI-GS(S)

	GLB		SEB		SwS	
	r	95% CIs	*r*	95% CIs	*r*	95%CIs
EX	0.58**	0.5; 0.65	-0.29**	-0.41; -0.21	-0,17**	-0.28; -0.06
CY	0,41**	0.31; 0.5	-0,28**	-0.38; -0.17	-0,39**	-0.48; -0.29
PE	-0,36**	-0.45; -0.26	0.51**	0.42; 0.59	0.41**	0.31; 0.50

Note: ** $p < 0.001$, GLB – general level of burnout or general burnout, SEB – self-efficacy beliefs, SwS – satisfaction with studies

cynicism subscales. The findings allow us to accept the assumed hypotheses for the construct validity of the adapted tool.

DISCUSSION

The aim of the present study was to investigate the value of a tool measuring burnout in Polish students. The analysis of the full version (16 items) supported their adequate three-factor structure, reliability, and validity – consistent with our hypothesis. Notably, the Cronbach's alpha for each scale exceeded 0.8, indicating strong internal consistency. A similar interpretation emerges from the correlations between variables, with the exception of one item (EX_5/MBI_6; "I feel burned out from my studies"), which has a higher correlation with Cynicism than Exhaustion. Removing this item enhanced the fit of the three-factor model. The same item lowered the reliability of the EX scale in the Polish version of the tool for measuring burnout in all professions (MBI-GS, Chirkowska-Smolak, Kleka, 2011), leading the authors of the adaptation to suggest its removal from the Polish version of the questionnaire.

Considering the fit coefficients of the three-factor model, our findings suggest the validity of the tool. While the results for the standard three-factor model were satisfactory, testing the modified version with the excluded item led to improved values. Comparable findings emerge in studies of other adaptations of the MBI-SS, but they involve the MBI_13 item (Hu, Schaufeli, 2009; Shin, Puig, Lee, J., Lee, J. H. & Lee, S. M., 2011; Tsubakita, Shimazaki, 2016, Faye-Dumanget, Carré, Le Borgne, Boudoukha, 2017). Additionally, we utilized MGCFA to assess the invariance of the MBI-SS across genders. The results indicate that the factorial structure of the MBI-SS remains consistent across both student samples.

LIMITATIONS OF THE STUDY

This study had its limitations. Primarily, our respondent sample was not entirely representative. The survey was primarily among students at public universities studying major subjects; thus, findings may not apply to students at technical or medical universities. A significant portion of the respondents were social science students, leading to a higher female representation. While our test of construct validity supported the adapted tool's value, concerns about the sample's representativeness persist (as noted above). Anticipated correlations between the surveys used would benefit from validation in a broader, more representative study.

A significant limitation is the lack of attention checks in our survey. The inclusion of attention checks has become an increasingly common practice in survey research to identify and potentially exclude respondents who may not be reading or processing questions thoroughly. Their absence has several implications: inattentive responses could introduce noise into the data, reducing the precision of our estimates and potentially limiting our confidence in interpreting results. To mitigate these concerns, we analyzed response patterns for straight-lining and other indications of inattentiveness, even though this is a more indirect measure than attention checks. We also considered the time participants took to complete the survey and found that no responses were completed in an unrealistically short timeframe. While these steps help alleviate some concerns, future iterations of this research should consider the incorporation of attention checks to further ensure data quality.

Further research should also consider criterion-relevance analysis and longitudinal studies for a more holistic understanding of academic burnout in students.

CONCLUSIONS

The MBI-SS questionnaire is invaluable for researchers, university educators, managers, and psychologists aiming to gauge the level of academic burnout in students. Our findings suggest that the Polish adaptation has good validity. The three-factor structure of the questionnaire was supported, comprising distinct factors: exhaustion, cynicism, and academic efficacy. Due to higher factor loadings for the MBI_6 item on another subscale (CY) rather than its original (EX), we recommend omitting this item from the Polish version. As such, we advocate for the use of the 15-item version of the scale (excluding MBI_6).

REFERENCES

Alarcon, G. M., Edwards, J. M., Menke, L. E. (2011). Student Burnout and Engagement: A Test of the Conservation of Resources Theory. *The Journal of Psychology* 145(3), 211-227. https://doi.org/10.1080/00223980.2011.555432

Baka, Ł. (2019). *Kopenhaski Kwestionariusz Psychospołeczny (COPSOQ II). Podręcznik do polskiej wersji narzędzia.* Centralny Instytut Ochrony Pracy – Państwowy Instytut Badawczy.

Beaton, D.E., Bombardier, C., Guillemin, F., Ferraz, M.B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine,* 25(24), 3186-3191. https://doi.org/10.1097/00007632-200012150-00014

Brenninkmeijer, V., Van Yperen, N. (2003). How to Conduct Research on Burnout: Advantages and Disadvantages of an Unidimensional Approach to Burnout, *Occupational and Environmental Medicine*, 60(1), 16-20. https://doi.org/10.1136%2Foem.60.suppl_1.i16

Campos, J. A. D. B., Carlotto, M. S., Marôco, J. (2012). Oldenburg Burnout Inventory-Student Version: Cultural Adaptation and Validation into Portuguese. *Psicologia: Reflexao e Critica*, 24(4), 709-718. https://doi.org/10.1590/S0102-79722012000400010

Campos, J. A. D. B., Carlotto, M. S., Marôco, J. (2013). Copenhagen Burnout Inventory-Student Version: Adaptation and Transcultural Validation for Portugal and Brazil. *Psicologia: Reflexao e Critica*, 26(1), 87-97. https://doi.org/10.1590/S0102-79722013000100010

Chirkowska-Smolak, T. (2012). *Psychologiczny model zaangażowania w pracę*. Wydawnictwo Naukowe UAM.

Chirkowska-Smolak, T., Grobelny, J. (2016). Konstrukcja i wstępna analiza psychometryczna Kwestionariusza Postrzeganego Stresu w Pracy (PSwP). *Czasopismo Psychologiczne*, 22(1), 131-139. http://dx.doi.org/10.14691/CPPJ.22.1.131

Chirkowska-Smolak, T., Kleka, P. (2011). The Maslach Burnout Inventory–General Survey: validation across different occupational groups in Poland. *Polish Psychological Bulletin*, 42(2), 86-94. http://dx.doi.org/10.2478/v10059-011-0014-x

Demerouti E., Bakker A.B., Nachreiner F., Schaufeli W.B. (2001). The job demands-resources model of burnout. *Journal of Applied*

Psychology, 86(3), 499–512. https://doi.org/10.1037/0021-9010.86.3.499

Esteve, E.B. (2008), *Well-being and Performance in Academic Settings: The Predicting Role of Self-efficacy* [unpublished doctoral dissertation]. Castellón de la Plana: Universidad Jaume I.

Faraci, P. (2018). Testing for the dimensionality of the Maslach Burnout Inventory (MBI) with a sample of high school teachers. Clinical Neuropsychiatry, 15(1), 50-59.

Faye-Dumanget, C., Carré, J., Le Borgne, M., Boudoukha, A. H. (2017). French validation of the Maslach Burnout Inventory-Students Survey (MBI-SS). *Journal of Evaluation in Clinical Practice*, 23(6), 1247–1251. https://doi.org/10.1111/jep.12771

Hu, Q.,Schaufeli, W. B. (2009). The factorial validity of the Maslach Burnout Inventory: Student survey in China. *Psychological Reports,* 105(2), 394–408. https://doi.org/10.2466/PR0.105.2.394-408

Judge, T.A., Bono, J.E., Locke, E.A. (2000). Personality and job satisfaction: The mediating role of job characteristics. Journal of Applied Psychology, 85(2), 237-49. http://dx.doi.org/10.1037/0021-9010.85.2.237

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika,* 39(1), 31–36. https://doi.org/10.1007/BF02291575

Kristensen, T.S., Borritz, M., Villadsen, E. and Christensen, K.B. (2005) The Copenhagen Burnout Inventory: A New Tool for the Assessment of Burnout. *Work and Stress*, 19, 192-207. http://dx.doi.org/10.1080/02678370500297720

Lee, R. T., Ashforth, B. E. (1996). A Meta-analytic Examination of the Correlates of the Three Dimensions of Job Burnout, *Journal of Applied Psychology,* 81(2), 123-133. http://dx.doi.org/10.1037/0021-9010.81.2.123

MacCallum, R. C., Browne, M. W., Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modelling. *Psychological Methods*, 1(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Maslach, C. (1993). Burnout: A Multidimensional Perspective. In: Schaufeli, W., Maslach, C., Marek, T. (eds.), *Professional Burnout: Recent Developments and Research* (19–32), Taylor & Francis: London. http://dx.doi.org/10.4324/9781315227979-3

Maslach, C., Jackson, S. E., Leiter, M. P. (1996). *Maslach burnout inventory manual (3rd ed.)*. Consulting Psychologists Press.

Maslach, C., Jackson, S. E., Leiter, M. P. (2017). *Maslach Burnout Inventory Manual*, *4th edition*. Mind Garden.

Maslach, C., Leiter, M. P. (2008). Early predictors of job burnout and engagement*, Journal of Applied Psychology*, 93(3), 498-512. http://dx.doi.org/10.1037/0021-9010.93.3.498

Maslach, C., Schaufeli, W. B., Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology,* 52(1), 397–422. https://doi.org/10.1146/annurev.psych.52.1.397

Pasikowski, T. (1996). Polska adaptacja kwestionariusza Maslach Burnout Inventory. In: Sęk, H. (eds.), *Wypalenie zawodowe: Przyczyny, mechanizmy, zapobieganie* (135-148). PWN. http://dx.doi.org/10.18559/SOEP.2018.3.2

Pejtersen, J. H., Kristensen, T. S., Borg, V., & Bjorner, J. B. (2009). The second version of the Copenhagen Psychosocial Questionnaire. *Scandinavian Journal of Public Health*.https://doi.org/10.1177/1403494809349858

Portoghese, I., Leiter, M. P., Maslach, C., Galletta, M., Porru, F., D'Aloja, E., Finco, G., Campagna, M. (2018). Measuring Burnout Among University Students: Factorial Validity, Invariance, and Latent Profiles of the Italian Version od the Maslach Burnout Inventory Student Survey (MBI-SS). *Frontiers in Psychology,* 9, article 2105. https://doi.org/10.3389/fpsyg.2018.02105

Revelle, W. (2021) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 2.1.6.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modelling. *Journal of Statistical Software*, 48(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Rostami, Z., Abedi, M. R., Schaufeli, W. B., Ahmadi, A., Sadeghi, A. H. (2014). The Psychometric Characteristics of Maslach Burnout Inventory Student Survey: A Study Students of Isfahan University. *Zahdan Journal of Research in Medical Sciences*, 16(9), 55-58. http://dx.doi.org/10.18637/jss.v048.i02

Salanova, M., Peiró, J. M., Schaufeli, W. B (2002). Self-efficacy Specificity and Burnout among Information Technology Workers: An Extension of the Job Demands-Control Model. *European Journal on Work and Organizational Psychology, 11*, 1-25. https://doi.org/10.1080/13594320143000735

Salgado, S., Au-Yong-Oliveira, M. (2021). Student burnout: A case study about a Portuguese public university. *Education Sciences*, 11(1), 1–35. https://doi.org/10.3390/educsci11010031

Schaufeli, W. B., Leiter, M. P., & Maslach, C. (2009). Burnout: 35 years of research and practice. *Career Development International, 14*(3), 204–220. https://doi.org/10.1108/13620430910966406

Schaufeli, W. B., Martinez, I., Pinto, A. M., Salanova, M., Bakker, A. B. (2002). Burnout and engagement in university students: A cross-national study. *Journal of Cross-Cultural Psychology,* 33(5), 464–481. https://doi.org/10.1177/0022022102033005003

Schaufeli, W. B., Taris, T. W. (2005). The conceptualization and measurement of burnout: Common ground and worlds apart. Work & Stress, 19, 256–262. https://doi.org/10.1080/02678370500385913

Simancas-Pallares, M. A., Fortich Mesa, N., González Martínez, F. D. (2017). Validity and internal consistency of the Maslach Burnout Inventory in dental students from Cartagena, Colombia. *Revista Colombiana de Psiquiatría (English Ed.)*, 46(2), 103–109. https://doi.org/10.1016/j.rcpeng.2017.05.009

Shin, H., Puig, A., Lee, J., Lee, J. H., Lee, S. M. (2011). Cultural validation of the Maslach Burnout Inventory for Korean students. *Asia Pacific Educational Review*, 12(4), 633–639. https://doi.org/10.1007/s12564-011-9164-y

Shirom, A. (2003). Job-related Burnout: A Review. In: Quick, J. C., Tetrick, L. (eds.), *Handbook of Occupational Health Psychology* (245-264). American Psychological Association. https://doi.org/10.1037/10474-012

Sörbom, D. (1989). Model modification. *Psychometrika*, 54(3), 371-384. https://doi.org/10.1007/BF02294623

Stawiarska, P. (2016). *Wypalenie zawodowe w perspektywie współczesnego świata*. Difin.

Tsubakita, T., Shimazaki, K. (2016). Constructing the Japanese version of the Maslach Burnout Inventory – Student Survey: Confirmatory factor analysis. *Japan Journal of Nursing Science*, 13(1), 183-188. https://doi.org/10.1111/jjns.12082

Turhan, D., Scheunemann, A., Schnettler, T., Bäulke, L., Thies, D. O., Dresel, M., Fries, S., Leutner, D., Wirth, J., Grunschel, C. (2021). Psychometric Properties of the German Short Version of the Maslach Burnout Inventory – Student Survey. *European Journal of Health Psychology*, 28(2), 45–58. https://doi.org/10.1027/2512-8442/a000067

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2):307-333. https://doi.org/10.2307/1912557

Wickramasinghe N. D., Dissanayake, D. S., Abeywardena, G. S. (2018). Validity and reliability of the Maslach Burnout Inventory-Student Survey in Sri Lanka. *BMC Psychology*, 6(1), article 52. https://doi.org/10.1186/s40359-018-0267-7

World Health Organization (WHO) (2016). Process of translation and adaptation of instruments. Available at: http://www.who.int/substance_abuse/research_tools/translation/en/ (26 January 2022).