



Research Article

Internal structure of an alternative measure of burnout: Study on the Slovenian adaptation of the Oldenburg Burnout Inventory (OLBI)



Nataša Sedlar^{a,*}, Lilijana Šprah^a, Sara Tement^b, Gregor Sočan^c

^a Sociomedical Institute, Scientific Research Centre of the Slovenian Academy of Sciences and Arts, Slovenia

^b Department of Psychology, Faculty of Arts, University of Maribor, Slovenia

^c Department of Psychology, University of Ljubljana, Slovenia

ARTICLE INFO

Article history:

Received 20 May 2014

Received in revised form 8 February 2015

Accepted 10 February 2015

Keywords:

Burnout

Oldenburg Burnout Inventory

Factor structure

Reliability

Psychometrics

ABSTRACT

This study evaluates the factorial validity and reliability of the Slovenian adaptation of the Oldenburg Burnout Inventory (OLBI) in a sample of 1436 Slovenian employees of various occupations. Confirmatory factor analyses were used to evaluate alternative structural models of OLBI, and reliability of variant scales was estimated. The results reveal a different structure of the Slovenian adaptation compared with the original one and a very notable difference in reliability between positively and negatively framed items. The results could be explained with a response bias or the specific nature of burnout and work engagement that OLBI promises to assess simultaneously. Therefore, we believe that the internal structure of the original inventory needs to be reconsidered.

© 2015 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Over the last decades occupational burnout gained an increased attention among professionals and researchers (for a review see Halbesleben & Buckley, 2004) due to its negative impact on employees' health, negative job attitudes and impaired organizational behavior (i.e., absenteeism, job turnover, presenteeism) (for a review see Schaufeli, 2003).

The most commonly-used definition of psychological burnout arises from Maslach and Jackson (1981), where burnout is defined as a syndrome consisting of three dimensions: emotional exhaustion, depersonalization and reduced personal accomplishment. Exhaustion occurs as a result of one's emotional demands. Depersonalization refers to a cynical, negative or detached response to care recipients/patients. Reduced personal accomplishment refers to a belief that one can no longer work effectively with clients/patients/care recipients. Following this conception authors developed the Maslach Burnout Inventory (MBI, Maslach & Jackson, 1981; Maslach, Jackson, & Leiter, 1996), which is currently the most widely used research instrument for burnout assessment.

Originally, the measure has been developed exclusively for use in human services professions (MBI-HSS). A second version of the MBI was developed for use in educational settings (MBI-ES). Due to increasing interest in burnout within occupations without a significant human service component, a third, general version of the MBI was developed (MBI-GS). There are several studies supporting the use of MBI for the assessment of burnout and its factorial validity across different occupations, languages and versions of MBI (for a recent meta-analysis of validation studies, see Worley, Vassar, Wheeler, & Barnes, 2008).

However, the construct's definition and measurement with MBI has drawn several criticisms. Some researchers (e.g. Kalliath, 2000) suggested that only the first two dimensions of emotional exhaustion and depersonalization should be included into the burnout model. Partly because the third dimension of personal accomplishment shows far less consistent relationships to some organizational outcomes (e.g. job satisfaction and organizational commitment; Lee & Ashforth, 1996) and probably could be more appropriately conceptualized as a personality trait similar to self-efficacy (e.g. Cordes & Dougherty, 1993). Furthermore, Demerouti, Bakker, Nachreiner, and Schaufeli (2001) pointed out that one-sided scales are inferior to scales that include mixed (both positively and negatively worded) items, because they can lead to artificial factor solutions in which positively and negatively worded items are likely to cluster.

To overcome these criticisms new inventories have been developed for the evaluation of the syndrome. One of the often used alternative burnout instruments, the Oldenburg Burnout Inventory

* Corresponding author at: Sociomedical Institute, Scientific Research Centre of the Slovenian Academy of Sciences and Arts, Novi trg 2, P.O. Box 306, 1001 Ljubljana, Slovenia. Tel.: +386 31447498.

E-mail addresses: natasa.sedlar@gmail.com (N. Sedlar), lilijana.sprah@guest.arnes.si (L. Šprah), sara.tement@um.si (S. Tement), gregor.socan@ff.uni-lj.si (G. Sočan).

(OLBI; Demerouti, Bakker, Vardakou, & Kantas, 2003), claims to solve both above-mentioned problems that are inherent to the MBI. It is based on a model similar to that of MBI, but employs only two dimensions (exhaustion and disengagement from work). Furthermore, both scales consist of mixed instead of only negative items, to mitigate the potential wording biases of the MBI. Contrary to the MBI that includes only the affective aspects of exhaustion, the OLBI also includes cognitive and physical aspects. According to authors this facilitates the application of OLBI to the workers that perform physical work or work with data. What is more, the disengagement dimension of OLBI refers to distancing oneself from one's work in general, thus exhibiting a cynical, negative attitude toward it, rather than only distancing oneself from people involved in work (e.g. coworkers, patients, clients), which is the case in the original MBI. Authors therefore argue that OLBI might be more generally applicable as compared to MBI, despite the fact that both instruments are suitable for any occupational group.

So far, several studies have confirmed factorial validity of the OLBI in different countries: Germany (Demerouti, Bakker, Nachreiner, & Ebbinghaus, 2002), the United States (Halbesleben & Demerouti, 2005), and Greece (Demerouti et al., 2003). The proposed two factor model demonstrated a relatively better fit to the data compared to alternative factor structures (unidimensional model, positive/negative wording model) in several occupational groups (human service, industrial, and transportation jobs).

On the other hand, some studies highlight potential limitations of the OLBI. For instance, Halbesleben (2003) noted that the fit statistics of two-factor models, obtained in his study, have been rather lower than regularly accepted levels. Although there was relatively more support for the two-factor structure (as compared to a unidimensional), the evidence for the construct validity of the OLBI was tentative only, due to the relatively poor fit of the tested models. The fit indices of the tested models were lower than regularly accepted levels proposed by Hu and Bentler (1999); e.g. RMSEA (<0.06); CFI (>0.95); TLI (>0.95) in other validation studies as well (e.g. Demerouti et al., 2003). What is more, studies in the United States (Halbesleben & Demerouti, 2005) and Greece (Demerouti et al., 2003), which have confirmed the convergent validity of the OLBI and MBI-GS, demonstrated that test-retest reliability of the OLBI dimensions for the time of 4 months was low (Halbesleben & Demerouti, 2005; $r_{\text{exhaustion}} = 0.51$, $r_{\text{disengagement}} = 0.34$).

In addition to that, the use of reversed items in measurement scales remains a controversial topic. Some authors, including Demerouti et al. (2003) recommend their use to reduce the potential effects of response pattern biases, while others advise against it, because the positive vs. negative framing of the items may act as a method factor obscuring the item structure of the measured trait (e.g. Weijters, Baumgartner, & Schillewaert, 2013). According to Weijters et al. (2013), there are three distinct mechanisms that could lead to method effects in response to reversed items: (a) acquiescence (preference for the positive or negative side of the rating scale), (b) careless responding (response that is not based on the content) and (c) confirmation bias (activation of beliefs that are consistent with the way in which the first item is stated). First two mechanisms encourage response inconsistencies between regular and reverse items, thus leading to correlated errors or the emergence of spurious factors. This is also in line with the notion of Podsakoff, MacKenzie, Lee, and Podsakoff (2003) that including reverse-coded items may produce artifactual response factors consisting exclusively of reverse-coded items. The third mechanism can lead to an upward or downward bias in respondent's scores, depending on the keying direction of the first item measuring focal construct. The method effects generated by these mechanisms may as well be present when all items are worded in the same direction, but are completely confounded with content variance and therefore undetectable, unless directly measured (Podsakoff

et al., 2003). Moreover, researchers have pointed out some other drawbacks of this approach. Including negatively and positively framed items may lead to interpretational problems, because positive and negative affective states have been shown to have different antecedents (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). Also, research on the structure of affect (Lloret & González-Romá, 2003) has demonstrated that low scores in positive items do not parallel high scores on negative items and vice versa (low scores on negative items do not parallel high scores on positive ones).

In light of the ambiguous findings regarding the factorial validity and item wording, we claim that is relevant to re-examine the psychometric properties of OLBI in an additional sample. Therefore, the aim of the presented study was to analyze the factorial structure and scale reliability of the Slovenian adaptation of OLBI. More particularly, we will compare the two-factor burnout model, consisting of two components of burnout (exhaustion and disengagement), with alternative structure models (unidimensional model, two-factor wording model (positive-negative wording), four factor model).

2. Method

2.1. Participants

The present study is based on two samples. As sample 2 is not very representative of the Slovenian workforce (i.e., lower educated and younger employees), another data collection method was simultaneously applied in order to secure greater heterogeneity and, in turn, generalizability of the findings. Based on a review by Wheeler, Shanine, Leon, and Whitman (2014) comparing student-recruited samples and organization-based samples, we also do not expect meaningful differences in the results obtained by the two samples.

Sample 1 was a student-recruited sample consisted of 1063 employees (58% female, 42% male), The most prevalent age group was 40–50 years (34%), 9% were younger than 20 years, 20% were aged between 20 and 30 years, 26% were aged between 30 and 40 and 10% were more than 50 years old. The educational structure was as follows: 30% obtained a university degree or higher, 21% completed a higher vocational school, 27% finished high school, others (22%) obtained a lower vocational education or basic (elementary) education. Approximately three quarters of the participants worked full-time (68%) and had a permanent long-term job contract (78%). Sample 2 was a heterogeneous sample obtained through five different organizations in health care, construction, and industrial work. Of the 373 employees, 48% were female and 52% were male. Twelve percent of the participants were younger than 20 years, 30% were aged between 20 and 30 years, 34% were aged between 30 and 40, others were aged 40 and 50 years. Twenty-three percent of this sample obtained a university degree or higher, 11% completed a higher vocational school, 31% finished high school, while others (31%) obtained a lower vocational education or basic (elementary) education. The vast majority of the employees worked full-time (98%) and had a long-term contract (92%).

The total sample consisted of 1436 Slovenian employees of various occupations, 749 of which were female and 687 were male. Eight percent of the participants were less than 20 years old, 24% were from 20 to 30 years old, 27% were from 30 to 40 years old, 33% were from 40 to 50 years old and 8% were more than 50 years old. Most of the participants completed either high school (28%), university (23%), higher vocational (19%) or vocational school (18%). The majority of participants worked with information (39%), 31% worked primarily with people and 28% worked primarily with things according to Things-Data-People taxonomy

(Fine & Cronshaw, 1999). They were employed in a wide variety of sectors: industry or manufacturing (28%), health care and social work (15%), education (10%), construction (7%), government, public administration and defence (5%), trade (5%), banking, financial services and insurance (4%), communication (4%), accommodation and food service (3%), arts, entertainment and recreation (3%), professional, scientific and technical activities (3%), transportation (2%) other or not defined (9%). The mean working experience was 16.4 years (SD = 10.7), the mean organizational tenure was 18.4 years (SD = 10.8). 82 percent of the sample worked under long-term and 16 percent under short-term contract.

2.2. Instruments

The Oldenburg Burnout Inventory (OLBI; Demerouti et al., 2001, 2003; Demerouti, Mostert, & Bakker, 2010) measures two dimensions of burnout: exhaustion and disengagement. Items are scored on a four-point scale from strongly agree (1) to strongly disagree (4). Each subscale includes four items that are positively framed and four items that are negatively framed. Positively framed items should be reverse-coded if one wants to assess burnout. The eight items of the exhaustion subscale refer to general feelings of emptiness, overtaxing from work, a strong need for rest, and a state of physical exhaustion. Example items are “After my work, I usually feel worn out and weary” and “After working, I have enough energy for my leisure activities” (reversed scoring). Disengagement subscale refers to distancing oneself from the object and the content of one’s work and to negative, cynical attitudes and behaviors toward one’s work in general. Example items are “It happens more and more often that I talk about my work in a negative way” and “I feel more and more engaged in my work” (reversed).

2.3. Procedure

2.3.1. The translation of the English OLBI in Slovenian language

The English version of OLBI obtained from authors (Demerouti et al., 2010) was translated into Slovenian language independently by three psychologists with good knowledge of both languages. Translations were compared and differences between them discussed and resolved. The final translation was back translated into English by a bilingual psychologist with a doctoral degree familiar with organizational psychology and by a professional interpreter with good knowledge of psychology. After comparing the back translations with the original inventory, some minor changes were made. With this preliminary form of OLBI, pilot data were collected in a small sample of researchers from a public Slovene research organization. Following the feedback, some additional minor changes were made.

2.3.2. Data collection

Data were collected in two ways: Sample 1 was obtained by psychology students as a part of the students’ requirements in an empirical research course. Students were instructed to distribute paper-and-pencil questionnaire to participants employed on regular terms (i.e., students working occasionally were not included). In order to secure data quality, students were given detailed instructions about the nature of the study and its relevance (Demerouti & Rispens, 2014). Measures were also taken to increase personal interest of the students (e.g. extra credit points). Sample 2 was obtained through five work organizations from different sectors (health, construction, industrial work). The data was collected as a part of the project “The Support Programme for Employers and Employees for Reducing Work-related Stress and Its Adverse Effects”. Contact persons at each organization were informed about the study and were asked to assist with data collection. The questionnaires were distributed before the start of an employee

training and returned immediately after completion. The approval of the local psychological ethics commission had been obtained prior to the study. Thus, confidentiality and anonymity were warranted.

2.4. Analyses

In total, there were 1.3% values missing (ranging from 0.3% to 3.3% across items). Missing data were imputed using the EM algorithm, which has been demonstrated to be an effective method of dealing with missing data (Graham, 2009), and all analyses were conducted using a total of 1436 participants.

To enable the equivocal interpretation of item difficulties, negatively framed items had been reversed before analysis. A higher score is thus related to a more positive valence.

Besides the standard descriptive statistics, we also computed an item difficulty index (IDI), defined as

$$IDI = 100 \left[\frac{M_j - \min_j}{\max_j - \min_j} \right]$$

where M_j stands for item mean score (reversed if necessary), and \max_j and \min_j stand for the maximum and minimum possible item score, respectively. IDI is therefore the item mean, interpolated within the possible range of item scores, and can be considered the generalization of the well-known item difficulty index for dichotomous items.

Confirmatory factor analysis was performed with the Mplus 6 program (Muthén & Muthén, 1998–2010). The WLSMV estimator, which is the default estimator for analyses with ordered categorical variables, was used. The following model fit indices were used besides the chi-square statistic (the approximate cut-off values and value-related references are in parentheses): RMSEA (<0.06; Hu & Bentler, 1999); CFI (>0.95, Hu & Bentler, 1999), TLI (>0.95; Hu & Bentler, 1999), WRMR (<0.90; Muthén, 1998–2004).

To evaluate the reliability of the scales implied by the tested models, we computed three internal consistency reliability coefficients: coefficient alpha, Guttman’s λ_2 and the greatest lower bound to reliability (glb; computed as proposed by Ten Berge, Snijders, & Zegers, 1981). All three coefficients are lower bounds to the true reliability in the sample, and their computation does not require unidimensionality of items to hold. We computed the scale scores using unit weights (1 or –1, respectively). In case of exploratory models, the allocation of items was based on the size of standardized loadings: each item was allocated to the scale related to factor with the highest loading, so that each item was allocated to one scale.

3. Results

This section consists of three parts. We begin with the presentation of the item descriptive statistics, followed by the evaluation of alternative structural models for OLBI by means of confirmatory factor analysis. Finally, we present the results of the reliability analysis for variant scales.

3.1. Descriptive statistics

Table 1 presents the descriptive statistics for the OLBI items. Items of the disengagement scale are listed first, followed by items of the exhaustion scale. The responses on all items ranged between 1 and 4.

A careful examination of item difficulties points to a possible acquiescence response bias. While the average difficulties for both scales were almost identical (mean IDI = 49), the average IDI for negatively framed items was 55.0, and the average IDI for

Table 1
Descriptive statistics for OLB items.

Item	M	IDI	SD	Skew.	Kurt.
D1 Interesting aspects	2.04	35	0.68	0.47	0.56
D2 Devaluation of work ^a	2.93	64	0.96	0.16	−0.96
D3 Mechanical execution ^a	2.52	51	0.91	0.01	−0.84
D4 Challenging	2.65	55	0.67	0.44	0.60
D5 Inner relationship ^a	2.35	45	0.93	0.11	−0.89
D6 Sick about work tasks ^a	2.62	54	0.83	−0.08	−0.61
D7 No other occupation	2.04	35	0.84	−0.57	−0.10
D8 More engaged	2.59	53	0.67	0.11	−0.15
E1 Tired before work ^a	2.54	51	0.71	−0.41	0.21
E2 Longer times for rest ^a	2.33	44	0.86	−0.14	−0.62
E3 Manageable tasks	2.70	57	0.70	0.34	0.07
E4 Emotionally drained ^a	2.63	54	0.87	0.03	−0.73
E5 Fit for leisure activities	2.93	64	0.73	0.17	−0.21
E6 Worn out ^a	1.90	30	0.86	0.13	−0.78
E7 Tolerable workload	2.38	46	0.59	0.48	1.72
E8 Feel energized	2.33	44	0.71	0.21	−0.11

Note: $N = 1436$ in all analyses; IDI, item difficulty index; D, disengagement scale, E, exhaustion scale.

^a Negatively framed item.

positively framed items was only 42.9. Taking into account that the negatively framed items have been reversed, this result indicates that the participants tended to prefer low ratings regardless of the item content. Of course, the possibility that this effect reflects substantive differences in item content cannot be ruled out.

3.2. Internal structure: dimensionality

Partly following Halbesleben and Demerouti (2005), we compared the proposed two-factor model, consisting of two components of burnout (exhaustion and disengagement), with four alternative models:

1. a unidimensional model, where all item correlations can be explained by a single common factor;
2. an alternative two-factor model that specifies factors based on positive and negative wording of items;
3. a reduced exhaustion–disengagement model, using only the negatively framed items. This model was the only model that was not stated in advance, but was based on the results of the analyses explained in the sequel;
4. a four factor model (as proposed by Qiao & Schaufeli, 2011), where exhaustion and disengagement were divided into positive and negative worded factors (positive exhaustion, negative exhaustion, etc.).

We should note that we did not compare a series of nested models, therefore we have not formally tested the differences in goodness-of-fit, so our model comparison is descriptive only. Because of the choice of the WLSMV estimator, which was optimal for categorical item-level data, we could not compute the information-based fit measures like AIC, which are useful in comparison of non-nested models, but require the use of a ML estimator.

Table 2 presents the goodness-of-fit indices used to assess the overall fit of the proposed models. The indices show that none of the models fits well. Nevertheless, the positive/negative model and the four factor model had a clearly better model fit than the remaining two a priori models. The fit of the proposed two factor model was in fact only marginally better than the fit of the single-factor model. On the other hand, the four factor model had the best fit indices, but the differences from the values pertaining to the positive/negative model were very small. The four factor model also had some serious problems. The factor covariance matrix was not positive definite, which means that the solution was not formally acceptable. Although this condition is directly related only to factor correlations, which are not meaningful, the presence

of such an improper solution makes the interpretation of other model parameter estimates (for instance, factor loadings) dubious at best. Additionally, the correlations between both positive factors and between both negative factors were very high (0.95 and 0.92, respectively), indicating bad discriminant validity of such factors. Finally, this model would imply very short scales, with only four items each, and consequently a relatively low reliability of the scales.

We also aimed to fit a four factor model where factors were defined in the same way as in both two-factor models, which means that each item was loaded on two factors. However, the estimation process of this model failed to converge. Such a model would also not be very satisfactory from a psychometric perspective, since it would imply scales with overlapping items.

We tested an ad hoc model, Reduced Negative Two-Factor Model, that used only the negatively framed 8 items, four items from each of the exhaustion and disengagement subscales. The reason for stating such a model was the general finding of the reliability analyses that negative items were more reliable than the positive items. Although the fit of this model was not entirely satisfactory either, the fit indices indicated a slightly better fit (and even notably better fit with regard to WRMR) than the positive/negative wording model, which was the best fitting a priori model.

3.3. Internal structure: reliability

The ultimate goal of the dimensionality analysis in the psychometric context is the establishment of useful scales. An important consideration in the process of selecting the optimal structural model is therefore the reliability of the scales implied by the model. Because only a single administration of the questionnaire was possible, only the method of internal consistency was a feasible method to estimate the scales' reliability. Following Nunnally and Bernstein (1994), p. 265, we evaluated coefficients larger than 0.80 as appropriate for group-level analyses, and coefficients larger than 0.90 as appropriate for individual diagnostics.

The values of the reliability coefficients are presented in Table 3. If all item responses were summed into a single scale score, its reliability would be appropriate at least for group-level analyses or screening purposes; however, a single score would not be acceptable from the dimensionality point of view. That is, although the single sum score would be quite highly reliable, it would not correspond to a single homogeneous construct, as follows from the poor values of goodness-of-fit indices. The originally proposed scales have a problematic dimensionality as well, and besides their

Table 2
Fit statistics for OLBI measurement model comparisons.

Model	χ^2	df	p	CFI	TLI	RMSEA	RMSEA 90% CI	WRMR
Unidimensional model	4153.3	104	<0.001	0.786	0.753	0.165	0.160–0.169	4.71
Positive/negative wording model	862.8	103	<0.001	0.960	0.953	0.072	0.067–0.076	2.15
Proposed two-factor model	4128.8	103	<0.001	0.787	0.752	0.165	0.161–0.169	4.69
Negative two-factor model	142.6	19	<0.001	0.992	0.988	0.067	0.057–0.078	1.08
Four factor model	789.4	98	<0.001	0.963	0.955	0.070	0.066–0.075	2.01

Note: N = 1436; RMSEA, root-mean-square error of approximation; RMSEA 90% CI, confidence interval for RMSEA; CFI, comparative fit index; TLI, Tucker Lewis index; WRMR, weighted root mean square residual $p(\text{RMSEA} < 0.05) < 0.001$ in all cases.

reliability coefficients are much lower than the one of the single scale, indicating a very limited psychometric usefulness.

The analyses of the remaining scale versions lead to general conclusion that only scales consisting from negatively framed items have satisfactory reliability. For instance, the glb value of all negative scales exceeds 0.80 even in cases when a scale consists of only four items. Scales consisting of positively framed items, on the other hand, had a relatively low or even outright unacceptable reliability.

4. Discussion

In the presented study we adapted the Oldenburg Burnout Inventory (OLBI) to Slovenian language and analyzed its internal structure in a sample of Slovenian employees.

Previous research (e.g. Demerouti et al., 2003; Halbesleben & Demerouti, 2005) suggests that a two-factor model with exhaustion and disengagement fits to the data better than an alternative two-factor model with positively and negatively framed items. However, the results of our study are not in line with these findings. The positive/negative model and the four factor model showed better model fit than the proposed two factor model and the single-factor model (Table 1). The latter two models appear unattractive from the psychometric viewpoint also because of low reliability and problematic dimensionality, which seems to be a problem of the four factor model, as well. The most suitable solution would therefore be the positive/negative wording model, which was the best fitting a priori model and has acceptable reliability. This finding agrees with Qiao and Schaufeli (2011) who reported poor fit of the proposed two factor model and the single-factor model, while positive/negative and four factor models fitted to the data better. Another option which was not originally considered by Demerouti et al. (2003) is negative two-factor model. This model includes only negatively framed items, which generally show higher reliability than the positive items, and fits the data slightly better than the positive/negative wording model. Nonetheless, this comparison should be taken with some precaution, since it is generally easier to achieve a good approximation to a low-dimensional structure with a smaller rather than a larger number of variables.

The superior fit to the data of the wording model (with positive and negative framing as factors) compared to the fit of the two-factor model (with exhaustion and disengagement) is somewhat

unexpected. As already mentioned before, the results should be taken with some precaution because we did not compare a series of nested models and our model comparison is descriptive only. We should also note that the p values related to the χ^2 statistic should be taken with some reservation because of the relative large sample size. There is also a very notable difference in reliability between positively and negatively framed items, the former having a relatively low or even unacceptable reliability, while the latter having a satisfactory reliability.

In our view, the findings indicate that the positively framed items either show a particular response bias or are measuring separate factors. The first argument is not in line with Demerouti et al. (2003) who suggest that existence of both positive and negative framed items in each factor forces respondents to reflect the content of items carefully, but is largely supported by other researchers. For example, Podsakoff et al. (2003) found that reverse-coding could be a source of common method bias, producing artifactual response factors consisting of reverse-coded items. This phenomenon was also reported by Schaufeli and Salanova (2007), who found that negatively worded scales (exhaustion, cynicism, and inefficacy beliefs) and positively worded scales (vigor, dedication, absorption, and efficacy beliefs) cluster together in two different second-order factors (burnout and engagement, respectively), which might be indicative of response bias. Also, Weijters et al. (2013) argued that the use of reversed items tend to lead to systematic differences in response to regular and reverse-keyed items and this method effect may be due to different reasons (acquiescence, careless responding or confirmation bias). Because the exhaustion and disengagement subscales include items that refer to their opposites (namely, vigor and dedication), we believe that especially the confirmation bias (i.e., activation of beliefs that are consistent with the way in which item is stated) could be of particular significance here. A previous experimental study addressing another work psychological variable (i.e., organizational commitment) has shown that situations producing higher degrees of fatigue are likely to result in an artificial positive/negative factor when using positively and negatively worded items (Merritt, 2012). As the OLBI was part of a very broad instrument and presented among the last in the composite questionnaire, the factorial structure of OLBI could also be affected by depleted mental resources of the participants. Future studies using the OLBI in its present form should, therefore, consider applying the questionnaire early in the study (Merritt, 2012).

The second possible explanation for the results obtained in our study pertains to the structure of the OLBI. Authors (Demerouti et al., 2003) assume that the dimensions of burnout (exhaustion, disengagement) and work engagement (vigor, dedication) are bipolar construct's representing each other's opposite. Thus, negatively framed items represent burnout, and the positively framed items represent engagement. Additional evidence for such reasoning can be found when reviewing item content. For instance, the positively framed OLBI-exhaustion item "When I work, I usually feel energized." seems fairly similar to the item "At my work, I feel bursting with energy" from the vigor dimension of the Utrecht Work Engagement Scale (Schaufeli, Bakker, & Salanova, 2006). Nevertheless, the results of Demerouti et al. (2010) show that

Table 3
Reliability estimates for variant scales.

Scales in the model	Alpha	λ_2	glb
Single scale	0.831	0.850	0.901
Exhaustion	0.733	0.764	0.823
Disengagement	0.713	0.741	0.809
Positive	0.698	0.709	0.759
Negative	0.880	0.885	0.906
Pos. exhaustion	0.552	0.557	0.589
Pos. Disengagement	0.599	0.604	0.646
Neg. exhaustion	0.826	0.829	0.840
Neg. disengagement	0.776	0.788	0.808

only one dimension of burnout, namely disengagement scale, contains items that represent a bipolar construct (i.e., questions on both ends of the disengagement-dedication continuum). On the contrary, the energy dimension of burnout (exhaustion subscale), which contains questions on both the ends of the exhaustion-vigor continuum, seems to represent two separated but highly related constructs. In a similar vein, our findings support the idea that the dimensions of burnout and work engagement included in the OLBI could measure different but highly related constructs, as positively framed items of OLBI do not load on the same factor as the negatively framed items. Furthermore, the examination of reliability coefficients reveals that negatively framed items (containing two factors of exhaustion and disengagement) appear to be satisfactory, while this does not hold true for positively framed items (vigor and engagement) that have low or even unacceptable reliability. Taken together, the results suggest that reporting different scores for the components of burnout and work engagement seems necessary since they could represent different constructs.

The results inhibit us from providing an answer to the question whether the poor fit of the proposed model is due to a response bias or the specific nature of burnout and work engagement that OLBI promises to assess simultaneously. A comprehensive procedure for testing the presence of various method effects was proposed by Weijters et al. (2013) after our data had already been collected and could not be used, because it requires some experimental manipulations in the data acquisition process. We nevertheless attempted to fit some simple models with method factors (for instance, models B, D and F from Weijters et al., 2013) which do not require such change. However, the fitting of these models failed because of convergence problems, caused probably by partial symmetry of method and substantive factors.

The analyses of OLBI reveal different structure on the Slovenian adaptation compared with the original one. The explanation that this could be due to an unsuccessful adaptation of the questionnaire cannot be ruled out but seems unlikely because the translation followed the standard protocol. Moreover, the systematic difference between positively and negatively framed items makes us believe that the internal structure of the original inventory needs to be reconsidered. The use of sophisticated models for testing method

effects, like the one proposed by Weijters et al. (2013), may be particularly beneficial in the future work.

We also need to note some potential limitations of our study. A first potential drawback concerns the reliance on self-reports. What is more, our study included a rather specific sample, which has not been randomly selected from the full range of possible occupations. This may raise concerns regarding the generalizability of results. Although the sample of participants represented a diverse number of employees from various workplace settings, it was predominantly restricted to employees of the industry or manufacturing, health care, social work and education. Moreover our sample was overrepresented by females, employees from 40 to 50 years old, and employees with either completed high school or university. Still, the previous research findings (Schutte, Toppin, Kalimo, & Schaufeli, 2000) indicate the same burnout structure in different subpopulations, so the influence of sample characteristics on the results may be expected to be of minor practical relevance.

The psychometric evaluation of the Slovenian translation of the OLBI reveals its different structure compared with the original one. On the basis of the results, we cannot recommend the use of the OLBI as a measure of burnout before the problem of method effect in response to reversed items is studied more systematically (for instance, as proposed by Weijters et al., 2013).

Acknowledgments

The presented study was a part of “The Support Programme for Employers and Employees for Reducing Work-related Stress and its Adverse Effects”, co-founded by the European Social Fund 2611-10-038312, EU (framework of the Operational Programme for Human Resources Development for the period 2007–2013) and Research Programme “Language, Memory and Politics of Representation”, co-founded by the Slovenian Research Agency P6-0347. *Conflict of interest statement:* The authors declare that there are no conflicts of interest.

Appendix

Table 4

Table 4
Item loadings and factor intercorrelations for the CFA solutions.

Item description	One factor	Original scales		Two factors		Negative scales		Four factors			
	F1	D	E	Positive/negative scales		D Neg	E Neg	D Pos	D Neg	E Pos	E Neg
				Pos	Neg						
Item loadings											
D1.Interesting.aspects	0.42	0.43		0.64				0.65			
D2.Devaluation.of.work ^a	−0.85	−0.87			0.86	0.87			0.88		
D3.Mechanical.execution ^a	−0.63	−0.64			0.66	0.67			0.66		
D4.Challenging	0.38	0.38		0.52				0.53			
D5.Inner.relationship ^a	−0.78	−0.79			0.80	0.82			0.81		
D6.Sick.about.work.tasks ^a	−0.74	−0.75			0.75	0.76			0.76		
D7.No.other.occupation	0.11	0.11		0.24				0.25			
D8.More.engaged	0.44	0.44		0.70				0.71			
E1.Tired.before.work ^a	−0.44		0.45		0.47		0.48				0.48
E2.Longer.times.for.rest ^a	−0.71		0.72		0.73		0.74				0.75
E3.Manageable.tasks	0.39		−0.39	0.57						0.57	
E4.Emotionally.drained ^a	−0.80		0.82		0.82		0.85				0.85
E5.Fit.for.leisure.activities	0.32		−0.33	0.46						0.46	
E6.Worn.out ^a	−0.78		0.80		0.80		0.83				0.82
E7.Tolerable.workload	0.26		−0.27	0.43						0.44	
E8.Feel.energized	0.47		−0.48	0.78						0.80	
Factor correlations											
Factor 2		−0.93		−0.36		0.92		−0.42			
Factor 3								0.95	−0.33		
Factor 4								−0.26	0.92	−0.36	

Note: Items loadings > |0.3| are shown in boldface. D, disengagement scale, E, exhaustion scale.

^a Negatively framed item.

References

- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*, 323–370.
- Cordes, C. L., & Dougherty, T. W. (1993). A review and an integration of research on job burnout. *Academy of Management Review, 18*, 621–656.
- Demerouti, E., Bakker, A. B., Nachreiner, F., & Ebbinghaus, M. (2002). From mental strain to burnout. *European Journal of Work and Organizational Psychology, 11*, 423–441.
- Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands resources model of burnout. *Journal of Applied Psychology, 86*, 499–512.
- Demerouti, E., Bakker, A. B., Vardakou, I., & Kantas, A. (2003). The convergent validity of two burnout instruments: A multitrait-multimethod analysis. *European Journal of Psychological Assessment, 18*, 296–307.
- Demerouti, E., Mostert, K., & Bakker, A. B. (2010). Burnout and work engagement: A thorough investigation of the independency of both constructs. *Journal of Occupational Health Psychology, 15*, 209–222.
- Demerouti, E., & Rispens, S. (2014). Improving the image of student-recruited samples: A commentary. *Journal of Occupational and Organizational Psychology, 87*, 34–41.
- Fine, S. A., & Cronshaw, S. F. (1999). *Functional job analysis: A foundation for human resources management*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576.
- Halbesleben, J. R. B. (2003). Assessing the construct validity of alternative measures of burnout: An investigation of the Oldenburg Burnout Inventory and the Utrecht Engagement Scale. In *Paper presented at the annual meeting of the Southern Management Association* Clearwater Beach, FL.
- Halbesleben, J. R. B., & Buckley, M. R. (2004). Burnout in organizational life. *Journal of Management, 30*, 859–879.
- Halbesleben, J. R. B., & Demerouti, E. (2005). The construct validity of an alternative measure of burnout: Investigating the English translation of the Oldenburg Burnout Inventory. *Work & Stress, 19*, 208–220.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
- Kalliath, T. J. (2000). A test of the Maslach Burnout Inventory in three samples of healthcare professionals. *Work & Stress, 14*, 35–51.
- 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*, 123–133.
- Lloret, S., & González-Romá, V. (2003). How do respondents construe ambiguous response formats of affect items? *Journal of Personality and Social Psychology, 85*, 956–968.
- Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Occupational Behavior, 2*, 99–113.
- Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *Maslach burnout inventory* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
- Merritt, S. M. (2012). The two-factor solution to Allen and Meyer's (1990) affective commitment scale: Effects of negatively worded items. *Journal of Business and Psychology, 27*, 421–436.
- Muthén, B. O. (1998–2004). *Mplus technical appendices*. Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (16th ed.). Los Angeles, CA: Muthén & Muthén.
- Nunnally, J., & Bernstein, I. (1994). *Psychometric theory*. New York: McGraw-Hill.
- Podsakoff, P. M., MacKenzie, S. M., Lee, J., & Podsakoff, N. P. (2003). Common method variance in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903.
- Qiao, H., & Schaufeli, W. B. (2011). The convergent validity of four burnout measures in a Chinese sample: A confirmatory factor-analytic approach. *Applied Psychology, 60*, 87–111.
- Schaufeli, W. B. (2003). Past performance and future perspectives of burnout research. *South African Journal of Industrial Psychology, 29*, 1–15.
- Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement, 66*, 701–716.
- Schaufeli, W. B., & Salanova, M. (2007). Efficacy or inefficacy, that is the question: Burnout and work engagement, and their relationship with efficacy beliefs. *Anxiety Stress & Coping, 20*, 177–196.
- Schutte, N., Toppinnen, S., Kalimo, R., & Schaufeli, W. (2000). The factorial validity of the Maslach Burnout Inventory-General Survey across occupational groups and nations. *Journal of Occupational and Organizational Psychology, 73*, 53–66.
- Ten Berge, J. M. F., Snijders, T. A. B., & Zegers, E. E. (1981). Computational aspects of the greatest lower bound to reliability and constrained minimum trace factor analysis. *Psychometrika, 46*, 357–366.
- Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods, 18*(3), 320–334.
- Wheeler, A. R., Shanine, K. K., Leon, M. R., & Whitman, M. V. (2014). Student-recruited samples in organizational research: A review, analysis, and guidelines for future research. *Journal of Occupational and Organizational Psychology, 87*, 1–26.
- Worley, J. A., Vassar, M., Wheeler, D. L., & Barnes, L. L. B. (2008). Factor structure of scores from the Maslach Burnout Inventory: A review and meta-analysis of 45 exploratory and confirmatory factor-analytic studies. *Educational and Psychological Measurement, 68*, 797–823.