Industrial Health 2014, **52**, 102–112

Original Article

The 9-item Bergen Burnout Inventory: Factorial Validity Across Organizations and Measurements of Longitudinal Data

Taru FELDT[1]*, Johanna RANTANEN[1], Katriina HYVÖNEN[1], Anne MÄKIKANGAS[1], Mari HUHTALA[1], Pia PIHLAJASAARI[1] and Ulla KINNUNEN[2]

[1] Department of Psychology, University of Jyväskylä, Finland
[2] School of Social Sciences and Humanities, University of Tampere, Finland

Received March 22, 2013 and accepted December 2, 2013
Published online in J-STAGE December 21, 2013

Abstract: The present study tested the factorial validity of the 9-item Bergen Burnout Inventory (BBI-9)[1)]. The BBI-9 is comprised of three core dimensions: (1) exhaustion at work; (2) cynicism toward the meaning of work; and (3) sense of inadequacy at work. The study further investigated whether the three-factor structure of the BBI-9 remains the same across different organizations (group invariance) and measurement time points (time invariance). The factorial group invariance was tested using a cross-sectional design with data pertaining to managers (n=742), and employees working in a bank (n=162), an engineering office (n=236), a public sector organization divided into three service areas: administration (n=102), education and culture (n=581), and social affairs and health (n=1,505). Factorial time invariance was tested using longitudinal data pertaining to managers, with three measurements over a four-year follow-up period. The confirmatory factor analysis revealed that the three-factor structure of the BBI-9 was invariant across cross-sectional samples. The factorial invariance was also supported across measurement times. To conclude, the factorial structure of the BBI-9 was found to remain the same regardless of the sample properties and measurement times.

Key words: Burnout, Bergen Burnout Inventory, Factor structure, Longitudinal study, Factor invariance

Introduction

The present study examined the factorial validity of the recently introduced 9-item Bergen Burnout Inventory (BBI-9)[1)]. The BBI-9 is designed to measure three core dimensions of burnout, namely emotional exhaustion (emotional component), cynicism (cognitive component), and sense of inadequacy (behavioural component)[1)]. In a previous cross-sectional study conducted among Finnish and Estonian managers, the three-factor structure of the BBI-9 scale was found to be supported[1)]. To gain further evidence for the usefulness of the BBI-9, we investigated the hypothesised three-factor structure of the BBI-9, using both heterogeneous cross-sectional organizational samples and a longitudinal sample of managers. These data sets allowed us to establish the factorial invariance of the BBI-9 across employees working in different organizations (group invariance) and across various measurement times (time invariance).

Burnout and its measurement

The burnout definition by Maslach and Leiter[2, 3)] has

*To whom correspondence should be addressed.
E-mail: taru.feldt@jyu.fi

received the most attention and has been widely cited in occupational health literature[4]. According to Maslach and Leiter[2, 3], burnout represents a persistent, work-related state of ill-being characterised by the dimensions of exhaustion, cynicism, and reduced professional efficacy. *Exhaustion* refers to the draining of emotional energy and feelings of chronic fatigue; *cynicism* describes having a distant and negative attitude toward one's job, and *reduced professional efficacy* refers to the belief that one is no longer effective in fulfilling one's job responsibilities[2, 3]. These dimensions are assumed to be negative consequences of chronic work stress[5–7].

Maslach Burnout Inventory (MBI)[8] was the first inventory that was designed to measure the three dimensions of burnout. Today, several versions of the MBI exist: the Human Services Survey (MBI-HSS), the Educators Survey (MBI-ES), and the General Survey (MBI-GS). Of these, the MBI-GS is the most widely used version of the MBI, as it was developed to measure job burnout in all professions independent of the vocational aspects[2, 4, 9]. It is noteworthy, however, that the MBI has been criticised for imprecision regarding the relationship between each of the three components to the antecedents and consequences of burnout, i.e., regarding the theoretical validity of burnout[7, 10–12]. Additionally, the professional efficacy dimension of the MBI has received criticism due to the positively worded items of the dimension[13] and because the dimension is suggested to reflect a personal characteristic rather than a burnout syndrome[14].

Like MBI-GS, the BBI was also developed to measure burnout in all occupations and it was based on the same theoretical three-dimensional burnout definition as MBI-GS. The original validation of the 25-item BBI was carried out in Norway[15]. Today, in Finland, its shortened 15-item version (the BBI-15)[16] is in wide use, particularly in the Occupational Health Services sector. The core idea in the development of the BBI-15 was to combine the best features of the MBI and the BBI-25 in the same measure, i.e., the target was to develop a burnout scale which would measure the three hypothesized components of burnout (like MBI) and, in addition, the total burnout score (like BBI-25)[16]. When shortening the BBI-25, Näätänen *et al.*[16] conducted series of confirmatory factor analyses and chose those items to the BBI-15 which had the best loadings to their respective factors. Also the content of the items were analysed; the items which described best the three core dimensions of burnout defined by Maslach, Schaufeli and Leiter[5] were chosen to the shortened scale.

According to the Näätänen *et al.*[16], the confirmatory factor analyses supported the expected three-factor structure of the BBI-15 among Finnish employees (total n>10,000). However, some minor error covariances between the items within the three factors emerged showing that the factor structure was not optimal. The burnout factors of the BBI-15 showed high intercorrelations, particularly for cynicism and inadequacy (reversed reduced professional efficacy) (0.91–0.93)[16]. Furthermore, the three dimensions and the total burnout score showed relatively high test-retest correlations (0.64–0.66) over a one-year follow-up study among 125 Finnish employees[16] lending support to the assumption that burnout is more a long-term reaction to stress than a short-term state. However, the factorial time invariance of the BBI-15 has not been studied in previous validation studies. The analysis of concurrent validity reported by Näätänen *et al.*[16] showed that the emotional exhaustion score of the BBI-15 correlated highly with the emotional exhaustion score of the MBI-GS (r=0.87) as well as did the cynicism scores (0.88). The third dimension of sense of inadequacy showed a weaker correlation (r=0.30) with the respective (reversed) MBI dimension. The high scores of the three dimensions of the BBI-15 have found to be correlated to different characters of work (e.g., low leadership support, poor organizational climate) as well as to other aspects of well-being (e.g., poor work ability, low life satisfaction)[16]. Today, the BBI-15 has also been used in empirical occupational health psychology studies[17–20].

The main difference between the BBI-15 and MBI-GS is that reduced professional efficacy items are worded differently. The MBI-GS measures the frequency of positive experiences of professional efficacy (reversed items in the total burnout score), whereas the BBI-15 estimates the sense of inadequacy at work. Thus, the criticism faced by the MBI[13] has been taken into account when developing the BBI-15; that is, the wording of the sense of inadequacy (labelled as professional efficacy in MBI-GS) items is negative and thus similar to the two other burnout dimensions of the BBI-15. In addition to the individual items, also the scales differ between the MBI-GS and BBI-15. The scale of the MBI-GS refers to the frequency of perceived feeling of exhaustion, cynicism, and reduced professional efficacy. In the scale of the BBI-15 employees evaluate the statements referring to exhaustion, cynicism, and inadequacy by using a Likert-type 6-point scale ranging from totally disagree to totally agree. Thus, the BBI measures more the intensity of the emotional exhaustion, cynicism, and inadequacy, whereas MBI-GS measures their frequency. The items of the BBI-15 are context-specific, i.e., all items refer to work.

The BBI-15 has been shortened very recently to improve its psychometrical properties. The study by Salmela-Aro *et al.*[1] tested the factorial validity of the BBI-15 among Finnish (n=742) and Estonian (n=414) managers. Their study showed that the expected three-factor structure of the BBI-15 (five items for each factor) did not show a flawless fit with the given managerial samples due to the cross-loadings and error covariances between the observed items. The study further indicated that the deletion of six items (with cross-loadings and error covariances) improved the model fit significantly, resulting in the shortened BBI-9 (three items for each factor). The three-factor structure of the BBI-9 provided an excellent fit with the given managerial samples. This structure was also invariant across Finnish and Estonian managerial samples; that is, the factor loadings, factor inter-correlations, and item error variances were the same across both managerial samples. This kind of measurement invariance is a fundamental prerequisite for its use when comparing different samples to each other, for example, regarding mean levels of burnout. In the present study, the factorial validity of BBI-9 was studied in addition to managers also in three organizations: 1) a public sector organization with three service areas, 2) a bank, and 3) an engineering company. These studied organizations differed both from managers as well as from each other regarding to the nature of employees' work.

The study by Salmela-Aro *et al.*[1] further showed that the concurrent validity of the BBI-9 was supported in cross-sectional studies conducted among Estonian and Finnish managers: When job stress was constructed by the effort–reward imbalance model[21, 22], high effort at work was associated with high work exhaustion, while high reward was related to low cynicism and low sense of inadequacy. A very recent longitudinal study by Feldt *et al.*[23] showed further that the highest BBI-9 scores over a four-year follow-up period were found among those managers who reported a high effort–reward imbalance and high personal overcommitment during the follow-up period. The same study indicated that an increase in the experiences of effort–reward imbalance over time was associated with increasing scores on the BBI-9 over time.

It is important to note, however, that the evidence for the factorial validity of the BBI-9 is so far limited to managerial samples only. It is therefore unclear whether the expected three-factor structure of the BBI-9 can be established regarding heterogeneous occupational samples (e.g., organizational samples) as well. Furthermore, the investigated manager samples in the study by Salmela-Aro *et al.*[1] were strongly male-dominated; consequently, firm conclusions cannot be drawn concerning the validity of the BBI-9 with respect to female-dominated samples. Finally, the factor structure of the BBI-9 has not been tested in longitudinal studies. It is therefore unknown whether the three-factor structure of the scale remains the same over time. The present study was designed to fill these voids in the previous research. These are fundamental issues to be examined before making firm recommendations regarding the wider use of the BBI-9 in the future.

The Present Study

Our main aim was to investigate the factorial validity of the BBI-9 inventory using data collected from employees working in three organizations (a public sector city organization with three service sectors, a bank, and an engineering company) and from young managers followed up at three measurement time points over a four-year follow-up period.

The used data differed from the previous factor analytic study of the BBI-9[1] providing us a good basis to test the factor structure of the BBI-9 in other kind of samples as well. Establishing the measurement invariance of the scale across different samples is essential to ensure that the investigated construct (three-dimensional burnout concept in this case) is fundamentally similar regardless of the sample differences. Particularly, the public sector city organization with three service areas is a valuable contribution of the present study because the previous BBI-9 study by Salmela-Aro *et al.*[1] focused solely on managers working primarily in the private sector. Employees working in a bank provided us a sample of employees working in the financial sector, a field which was not included in the Finnish sample of managers in the study by Salmela-Aro *et al.*[1] and was in very minor role (6%) in a sample of Estonian managers[1]. The engineering company represented technical designing, a field, which was not included in the previous factor analytic study of the BBI-9[1]. Finally, our longitudinal study with three measurement times over a four-year follow-up time provided us the possibility to investigate the stability of the factor structure of the BBI-9 across time. This issue could not be addressed in the cross-sectional study by Salmela-Aro *et al*[1].

To summarise our factor analytic aims, we stated three main research questions to the present study:

1) Is the three-factor structure of the BBI-9 (i.e., three correlated factors of exhaustion, cynicism, and inadequacy) established in the study by Salmela-Aro *et al.*[1] valid

also in the present samples?

2) Is the three-factor structure of the BBI-9 invariant across investigated samples, i.e., does the BBI-9 meet the criteria of factorial group invariance?

3) Is the three-factor structure of the BBI-9 invariant across three measurement times, i.e., does the BBI-9 meet the criteria of factorial time invariance?

Finally, our secondary aim was to provide descriptive information about the mean scores of the BBI-9 across different organizational samples as well as across measurement times in the longitudinal data of young managers. There exist no previous investigations about the mean scores of the BBI-9 factors across different samples or measurement times over time. To summarise our descriptive aims, we stated the following two research questions:

4) Are there mean differences in the BBI-9 scores (i.e., the total BBI-9 score, sub-scores of exhaustion, cynicism, and inadequacy) between the investigated organizational samples?

5) Are there mean changes in the BBI-9 scores (i.e., the total BBI-9 score, sub-scores of exhaustion, cynicism, and inadequacy) across the three measurement times in a longitudinal study conducted among young managers?

Method

Participants and procedure

The three organizations examined in this study were 1) a public sector organization with three service areas, 2) a bank, and 3) an engineering company. In all organizations, the data were collected from the organization's employees using an internet-based questionnaire, in May 2011. Before the study was executed, the content of the research project and the procedure were agreed on with the organizations' management. All the employees were informed using an electronic bulletin describing the project and its purpose, which was available in the organizations' intranet a week before the data were collected. The following week, a link to the questionnaire, together with a cover letter (stating that the participation in the study was voluntary and confidential), was sent to the employees' e-mail addresses through the organizations' intranet.

Public sector organization

When the data collection was performed (May 2011), the public sector organization included altogether 8,366 employees who worked in four units (service areas): 1) administration (n=268, 3%), 2) urban design and business activities (n=1,766, 21%), 3) social affairs and health (n=4,359, 52%), and 4) education and culture (n=1,973, 23%). Altogether 3,402 respondents began answering the Internet-based questionnaire (all of whom did not complete all items), yielding a response rate of 40.7% (3,402 of 8,366). We included in the present study only those respondents who had answered all BBI items and who worked in administration, social affairs and health, and education and culture. Those respondents working in urban design and business activities were left out of this study because of the low response rate of this sector (response rate for the BBI-9 was 16%).

Administration. Altogether 102 of 268 employees working in administration filled the BBI-9 (response rate for the BBI-9 was 38%). The majority of them were women (n=83, 81%), and the largest age group consisted of 51–55-yr-olds (25%). The most typical educational level in this group was that of a secondary education (39%). The majority had a permanent employment contract (94%) and worked a regular day shift (93%). The participants from the administrative centre typically worked 38 h per week (SD=5.6), and 13% had managerial tasks. Typical vocational roles included specialists from different areas, designers, managers, and secretaries.

Social affairs and health. Of the employees working in the social affairs and health sector, 35% (n=1,505) responded to the BBI-9. The great majority of them (94%) were women. The most common age group was comprised of 51–55-yr-olds (18%), and the most typical educational level was that of a secondary education (36%). Of these participants, 85% had a permanent employment contract and 60% worked in a regular day shift, on average 38 h per week (SD=7.3). Only 8% had managerial tasks. In the social affairs and health sector, the most common vocational groups were nurses and nursery school teachers.

Education and culture. Altogether 581 participants from education and culture responded the BBI-9 items (29%). Of the, 74% were women. Of the age groups, the group comprised of 46–50-yr-olds was the largest (18%), and half of the participants had a university degree (50%). The majority of these participants had a permanent employment contract (84%) and worked a regular day shift (83%). The average working hours per week were 36 (SD=6.8) and 10% had managerial tasks. The largest occupational groups were teachers and teaching assistants.

Bank

Altogether 274 employees worked at the bank, in May 2011. Of these participants, 192 (70%) responded to the questionnaire. The target group of the present study con-

sisted of those 162 employees who had filled in all of the BBI items (the BBI-9 response rate was 59%). The majority (80%) of these were women, and two of the largest age groups were 51–55-yr-olds (19%) and 46–50-yr-olds (19%). Regarding their educational background, 43% had a college degree and 25% a professional college or lower university degree. The majority (93%) had a permanent employment contract. The participants typically worked 37 h per week (SD=8.4). Only a minority of the participants (13%) had managerial tasks.

Engineering company

The engineering company had 536 employees and 300 responded to the questionnaire (56%). Of the respondents, 236 had answered the BBI items (the BBI-9 response rate was 44%). The great majority of the participants consisted of men (78%). Two of the largest age groups were 26–30-yr-olds (18%) and 31–35-yr-olds (16%). Regarding the educational level, 53% had a professional college or lower university degree, and 25% had completed a university degree. The majority had a permanent employment contract (98%). The participants worked 39 h per week (SD=4.9), on average. Only a minority of the participants (22%) had managerial tasks. Most of these employees had design tasks.

A three-wave, four-year follow-up study of managers

The longitudinal sample of the present study included young professionals who, at the study baseline, worked in management or leadership positions. The participants had been followed up over four years at three measurement time points: 2006 (Time 1, T1), 2008 (Time 2, T2), and 2010 (Time 3, T3). The sample was taken from the membership registers of two Finnish national labour unions (Trade Union Pro, and the Union of Professional Engineers). The original sample included 1,904 union members who were all younger than 36 yr, and whose professional title referred to a management or leadership position.

In the first study phase, in 2006, 933 of the 1,904 postal questionnaires sent were returned. Of the respondents, 174 returned the attached paper slip, indicating that they did not fulfill the study criteria, i.e., they were not in a management or leadership position or in employment (e.g., they were studying full-time, or were unemployed for over 3 months). Therefore, these respondents were excluded from the baseline sample at T1, yielding a response rate of 44% (n=759)[17, 19]. In this study, we included those 742 respondents who answered the BBI items. The average age of these participants at T1 was 31 yr (range 23–36 yr, SD=3.2 yr) and a large majority of the participants were men (86%). The majority of participants were engineers (67%) and other participants were technicians (6%) or had other professional qualifications (25%). Only 2% of the participants had no professional qualification. Of the participants, 8% were working in upper, 49% in middle, and 43% in lower management. Of the participants, 93% had a permanent employment contract and they worked, on average, 43 h per week (SD=7.7).

At T2, in 2008, the follow-up questionnaires were sent to the 621 participants who had not declined further participation in the study. Altogether 433 participants returned the questionnaire completed, resulting in a response rate of 70%. Compared to the baseline sample (n=759), 57% of the participants were still participating in the study at T2. Of the participants, 421 had responded to the BBI items at T2. Of the 421 participants, 84% were men, and 10% worked in upper, 46% in middle, and 26% in lower management. Eighteen percent of the participants were no longer in a management position or did not report their managerial level. Also, 28% of the participants reported that they changed their job between T1 and T2.

At T3 in 2010, the questionnaires were sent to those who had answered at the study baseline and had not refused to be contacted again in later study phases (n=595). Altogether 380 participants returned the questionnaire completed, yielding a response rate of 64%. Of the baseline sample (n=759), 50% of the participants still participated in the study at T3. In total, 369 participants had responded to the BBI items at T3. Of the 369 participants, 83% were men, and 13% worked in upper, 46% in middle, and 18% in lower management. At T3, 23% of the participants were no longer in a management position or did not report their managerial level. Between T2 and T3, 17% of the participants changed jobs.

The attrition analysis performed at T1 has shown that the respondents did not differ in gender from the non-respondents[17]. Data on the non-respondents' age was only available for the members of the Union of Salaried Employees; these respondents did not differ from the non-respondents in age[17]. The attrition analyses of the longitudinal data have shown that those subjects who participated in the study at T2 and T3 did not differ in gender, age or managerial level from those who left the study after T1[23].

Measures

The BBI-9 consists of 9 items with three subscales measuring work exhaustion (three items), cynicism (three items), and sense of inadequacy (three items)[1]. The origi-

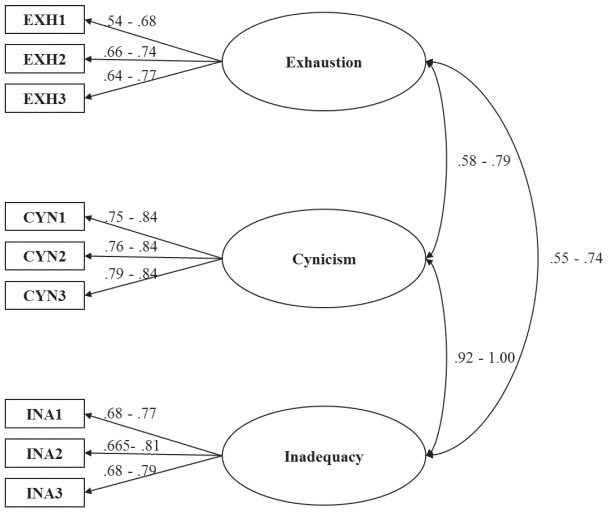

Fig. 1. The ranges of standardised factor loadings and factor covariances for the three-factor model of the BBI-9, across the six cross-sectional study samples. The observed items are numbered according to their item numbers presented in the study by Salmela-Aro *et al.*[1)]

nal items of the BBI-9 are available in the appendix of the study by Salmela-Aro *et al*[1)].

Analytic strategy

We used confirmatory factor analysis (CFA) and structural equation modelling (SEM) in investigating the factorial validity of the BBI-9. The statistical analyses were performed with the Mplus statistical package[24)], using the missing data method and robust maximum likelihood (MLR) as the method of estimation, since these are robust to non-normality in the observed variables.

We first tested the factorial group invariance of the hypothesised three-factor structure of the BBI-9, using six cross-sectional samples (the three service sectors of the public sector city organisation, the bank, the engineering company, and the sample of managers collected at T1). At first, the goodness-of-fit of the three-factor model was tested separately for each sample. Then, the factorial group invariance across samples was tested by comparing the freely estimated three-factor model (factor loadings allowed to vary across samples) to the constrained three-factor model (factor loadings set to be equal across samples)[25)]. The invariance in the factor loadings was evaluated by using the Satorra–Bentler scaled χ^2 difference test[26)]: If the test produced a non-significant loss of fit for the constrained model compared to the freely estimated model, then the invariance assumption was supported.

In the final phase of our analyses, we tested the factorial time invariance of the BBI-9 with respect to our three-wave, four-year follow-up study on managers. To do this, we estimated a longitudinal factor analysis model by connecting the three-factor models estimated at each measurement time point, using structural equations correlating the parallel factors across time. Again, the invariance of the factor loadings across time was tested by using the Satorra–Bentler scaled χ^2 difference test.

The goodness-of-fit of each estimated CFA and SEM model was evaluated using the following three absolute goodness-of fit indices: (1) χ^2 test, (2) Root Mean Square Error of Approximation (RMSEA) and (3) Standardised Root Mean Square Residual (SRMR). A non-significant χ^2 test indicates a good fit[27)], as do RMSEA and SRMR with values of 0.05 or less, whereas values of 0.06–0.08 indicate a reasonable fit, and values ≥0.10 a poor fit[28)]. Because the χ^2 test is sensitive to sample size, the use of relative goodness-of fit indices is also strongly recommended in the case of large sample sizes [29)]. Consequently, the following relative goodness-of-fit indices were also used to evaluate model fit: (1) Comparative Fit Index (*CFI*) and (2) Tucker-Lewis Index (*TLI*) with values ≥0.90 indicating a good fit[29)].

To respond to the research question 4 (mean differences in burnout scores between organizations), we used analysis of covariance (ANCOVA). As organizations differed from each other in regards to employees' gender, age, and educational background, we controlled these variables from the analyses. In organizational samples, age was classified into nine categories dividing age into five-year periods starting from under 25 yr and ending to the age group of over 60 yr Education was classified into six categories (1=basic education, 6=university degree). Repeated measures ANCOVA was used in analysing mean changes in burnout scores across time (research question 5) in the sample of managers. Again, gender, age (continuous variable), and education (1=basic education and short vocational courses, 5=university degree in engineering) were controlled for. All ANCOVAs were calculated by using the SPSS 16 and PASW Statistics 18 programs.

Results

The goodness-of-fit of the three-factor structure of the BBI-9 in cross-sectional samples

Table 1 shows the fit indices for the hypothesised three-factor model of the BBI-9 for each cross-sectional sample of the study. As seen, the three-factor model showed a very good fit with the given samples. In all samples, the

Table 1. Goodness-of-fit criteria for the correlated three-factor BBI-9 model, per sample

Sample	*n*	χ^2	*df*	*p*	*RMSEA*	*SRMR*	*CFI*	*TLI*
Public sector city organisation								
Administration	102	33.75	24	0.089	0.06	0.06	0.97	0.95
Education and culture	581	71.88	24	0.000	0.06	0.04	0.97	0.95
Social affairs and health	1,505	79.10	24	0.000	0.04	0.02	0.99	0.98
Bank	162	19.41	24	0.730	0.00	0.03	1.00	1.02
Engineering company	236	35.24	24	0.065	0.05	0.04	0.98	0.98
Managers at Time 1	742	47.90	24	0.003	0.04	0.03	0.99	0.98

RMSEA = Root Mean Square Error of Approximation, SRMR = Standardised Root Mean Square Residual, CFI = Comparative Fit Index, TLI = Tucker–Lewis Index

CFI and TLI values were excellent (i.e., above or equal to 0.95) and RMSEA and SRMR values varied from good (i.e., below or equal to 0.05) to reasonable (i.e., 0.06).

The factorial group invariance of the BBI-9

MacCallum, Browne, and Cai[30] have stated that, with very large samples, such as here (n=3,328 in the multi-group model containing all cross-sectional samples), a significant loss-of-fit for the invariant model is expected because the χ^2 difference test is based on the standard *null hypothesis* of the *zero difference* between the nested models. In the case of large samples with strong statistical power, this inevitably leads to the rejection of the null hypothesis, although, in reality, the difference in fit between the compared nested models might be *small* instead of *zero*. Thus, following the recommendation by MacCallum *et al.*[30], we relied on the fact that the RMSEA values between our freely estimated and constrained models did not show a significant difference but instead were also supporting the equality constraints in the factor loadings, that is, the factorial group invariance of the three-factor BBI-9 model. The range of standardised factor loadings and factor covariances for this model are depicted in Fig. 1.

The factorial time invariance of the BBI-9

Using the three-wave, four-year follow-up data on managers, we further investigated the factorial validity of the BBI-9 across time. As seen in Table 1, the three-factor model showed a good fit with the manager sample from the study baseline in 2006 (T1). The invariance in the factor loadings of the three-factor model during the follow-up time was tested by comparing the freely estimated model (factor loadings allowed to vary across the measurement times to the constrained model (factor loadings set to be equal across the three measurement times). The fit indices for the freely estimated three-factor model were good: $\chi^2(282)=458.89$, $p=0.000$, $RMSEA=0.03$, $SRMR=0.05$, $CFI=0.97$, $TLI=0.96$. The corresponding indices for the constrained model were also good: $\chi^2(294)=470.23$, $p=0.000$, $RMSEA=0.03$, $SRMR=0.05$, $CFI=0.97$, $TLI=0.96$. The Satorra–Bentler scaled χ^2 difference test produced a non-significant loss-of-fit for the time-invariant three-factor model: $\chi^2(12)=11.82$, $p=0.461$; and consequently, the time invariance assumption of the BBI-9 received support. The β coefficients between T1 and T2, and between T2 and T3, were respectively: 0.50 and 0.70 for exhaustion, 0.46 and 0.59 for cynicism, and 0.42 and 0.51 for inadequacy; indicating a relatively high rank-order stability regarding each dimension.

Descriptive information of the BBI-9 scores

Table 2 shows descriptive statistics (means and standard deviations) and the scale's internal consistency reliability coefficients (the Cronbach's alphas) for the total score regarding burnout according to the BBI-9, as well as for the scores regarding the dimensions of exhaustion, cynicism, and sense of inadequacy. As seen, the reliabilities of all the scores were acceptable. The results of the ANCOVA (gender, age, and education controlled for) showed that emotional exhaustion was highest in the social affairs and health sector whereas the cynicism score was highest in the engineering company. Inadequacy was highest in the bank and engineering company. Repeated measures ANCOVAs (gender, age, and education controlled for) across time among managers did not reveal significant mean changes regarding the total burnout score and its three sub-scores.

Discussion

This study was designed to examine the factorial validity of the recently introduced 9-item BBI[1], a brief scale

Table 2. Means (M), Standard deviations (SD), and Cronbach's alpha coefficients (α) for the BBI-9 scores, per sample

		Burnout (9 items)		Exhaustion (3 items)		Cynicism (3 items)		Inadequacy (3 items)	
Sample	*n*	*M* (*SD*)	α	*M* (*SD*)	α	*M* (*SD*)	α	*M* (*SD*)	α
Organisational samples									
1 Administration	102	2.64 (0.94)	0.88	2.69 (1.01)	0.74	2.35 (1.01)	0.84	2.87 (1.26)	0.81
2 Education and culture	581	2.49 (0.91)	0.87	2.67 (1.00)	0.70	2.22 (1.05)	0.84	2.60 (1.21)	0.77
3 Social affairs and health	1,505	2.56 (1.02)	0.90	2.79 (1.07)	0.72	2.26 (1.19)	0.87	2.66 (1.29)	0.83
4 Bank	162	2.68 (0.93)	0.87	2.64 (0.97)	0.67	2.43 (1.13)	0.84	2.98 (1.19)	0.76
5 Engineering company	236	2.76 (0.94)	0.88	2.59 (0.97)	0.76	2.70 (1.13)	0.84	2.99 (1.20)	0.78
Managers									
Time 1 (T1, in 2006)	742	2.52 (0.85)	0.85	2.85 (1.05)	0.70	2.27 (1.02)	0.82	2.44 (1.02)	0.71
Time 2 (T2, in 2008)	421	2.59 (0.96)	0.88	2.88 (1.23)	0.75	2.36 (1.09)	0.83	2.52 (1.21)	0.82
Time 3 (T3, in 2010)	369	2.66 (0.97)	0.88	2.85 (1.07)	0.72	2.46 (1.14)	0.85	2.66 (1.22)	0.80

ANCOVA for Burnout across groups: F (4) = 2.975, p=0.018; no significant group differences at p<0.05 level. ANCOVA for Exhaustion across groups: F (4) = 4.483, p=0.001 significant group differences: 2, 5 < 3. ANCOVA for Cynicism across groups: F (4) = 4.164, p=0.002; significant group differences: 2 < 5. ANCOVA for Inadequacy across groups: F (4) = 5.049, p=0.000; significant group differences: 2 < 4, 5; 3 < 4. Repeated measures ANCOVA across time for Burnout among managers: F (2,288) = 0.647, p=0.524. Repeated measures ANCOVA across time for Exhaustion among managers: F (2, 288) = 0.114, p=0.892. Repeated measures ANCOVA across time for Cynicism among managers: F (2, 288) = 1.323, p=0.268. Repeated measures ANCOVA across time for Inadequacy among managers: F (2, 288) = 0.343, p=0.710

that is developed to measure three core dimensions of burnout: exhaustion (3 items), cynicism (3 items), and sense of inadequacy (3 items). The group invariance of the hypothesised three-factor structure of the BBI-9 was tested across six cross-sectional samples, including a public sector organisation with three service sectors, a bank, an engineering company, and a sample of managers working in different organisations (at T1). The time invariance of the scale was investigated by using three-wave follow-up data collected over a four-year period, pertaining to managers.

The results indicated that the three-factor structure of the BBI-9 fitted well with all six samples. Furthermore, the three-factor structure of the BBI-9 turned out to be invariant across six cross-sectional samples (i.e., five organizational samples and managerial sample at T1). The finding of factorial invariance indicates that the factor loadings of the three-factor model of the BBI-9 did not vary significantly across samples, and, consequently, the scale items and their interrelations were perceived to be the same regardless of the sample properties. Thus, the mean differences that occurred between the samples in the BBI-9 scores cannot be explained by the structural variation of the scale for the investigated groups.

Our longitudinal SEM analysis provided more support for the factorial invariance of the BBI-9: The three-factor structure remained the same across the three measurement times points over the four-year follow-up period among investigated managerial sample. Thus, the BBI-9 is also a valid measure when used in longitudinal studies, and no mean differences due to instability in the factor structure of the scale occur in the BBI-9 scores over time.

Despite the strong evidence of factorial invariance in the BBI-9 across cross-sectional samples, it must be noted, however, that the inter-correlations between cynicism and inadequacy factors were very high indicating that these dimensions had little variance of their own. Exhaustion, in turn, showed weaker correlations to cynicism and inadequacy, indicating that it has a unique character. In this respect, the BBI-9 differs from the MBI-GS. In the MBI-GS, exhaustion and cynicism show high inter-correlation whereas reduced professional efficacy correlates relatively weakly to exhaustion and cynicism scores[4, 31]. These correlative differences between the BBI-9 and the MBI-GS could be explained by subtle differences in nuance between the cynicism scales. Cynicism in the MBI-GS is an interpersonal dimension (i.e., cold and impersonal relationships towards colleagues, clients, etc.). However, this is not the case of the BBI-9 in which cynicism refers more respondents' cynical feelings towards him/herself. This feature of cynicism could also explain the strong association between cynicism and inadequacy dimensions in the BBI-9.

Thus, when using the BBI-9, it should be remembered that the cynicism and inadequacy are highly interrelated dimensions. However, this does not mean that their intensity would be similar. As the results of the present

study showed there were significant mean differences in the BBI-9 scales between the investigated samples, which partly promote the use of separate subscales (Table 2). Thus, despite the strong inter-correlation between cynicism and inadequacy, their intensity does not always go hand in hand.

In general, the observed means of the BBI-9 scales (Table 2) showed that the participants scored rather low on burnout and its three dimensions of exhaustion, cynicism, and sense of inadequacy. However, some interesting mean differences between five organizational samples were observed. Exhaustion was particularly high in the social affairs and health sector of the public sector organization, a finding which is not surprising in the light of the recent public discussion in Finland concerning the reduced resources of social and health sector organizations. The financial cut downs were put into practice also in the investigated social affairs and health organization at the time when our data was collected. It is thus possible that the reduced financial resources led to higher work load among employees in the social and health sector, which can explain their higher exhaustion scores.

Employees in the engineering company reported the highest levels of cynicism and inadequacy. This finding can be partially explained by the economic recession in domestic and international markets, which had led to temporary lay-offs in the company during the year of the data collection (in 2011). Lay-offs have previously been found to be related to increasing levels of cynicism[32]. Employees working for the bank also reported relatively high levels of inadequacy. Again, the challenging financial market, particularly in the European Union, may have contributed to the feelings of inadequacy among banking staff, as their job demanded, for instance, advising customers and selling various investment products to them despite an uncertain market climate.

Strengths, limitations and future directions

The strengths of the present study were the use of large and heterogeneous samples in investigating the construct validity of the BBI-9. Further, applying three measurement times to our longitudinal study made a strong contribution, enabling us to empirically confirm the structural stability of the BBI-9 over time. However, despite these strengths, our study also has some limitations that reduce the generalisability of the results. First, the response rates of the present study were quite low (from 29 to 59%) which limits the generalisability of the present results. Differences in response rates in electronic surveys may be explained by the different nature of respondents' jobs in the investigated organisations. For example, employees in education and culture (e.g., teachers, teaching assistants) do not use computer in their work as much as employees working in a bank. So, it is reasonable to expect that the survey sent to email has not reached all employees in education and culture in time.

Second, our samples were all Finnish, and therefore firm conclusions cannot be made concerning the structural validity of the BBI-9 with regard to samples gathered in other countries. More cross-cultural evidence is still needed to establish the structure of the BBI-9 universally, since research has thus far been limited to Finnish samples and Estonian managers[1]. Finally, an important task for future research is to investigate the other validation types for the BBI-9, such as its criterion validity (e.g., associations with predictors and consequences of burnout). Related to this, the BBI-9 should be investigated in relation to other three-dimensional burnout measures such as MBI[8] and Shirom-Melamed Burnout Measure (SMBM)[33].

The rank-order stability of the BBI-9 scores should also be investigated in future inquiries. Following the core idea of the burnout concept, the BBI-9 should measure burnout more as a long-term rather than a short-term state and, consequently, the test-retest stabilities of the scores should be relatively high over time. Our longitudinal SEM analyses provided some knowledge on the rank-order stability of the BBI-9 scores, i.e., the degree to which the relative order of participants according to their BBI-9 scores is maintained over time. The stability coefficient for exhaustion turned out to be moderate between T1 and T2, and relatively high between T2 and T3. The corresponding stabilities for cynicism and inadequacy varied from relatively low to moderate. However, it must be remembered that the investigated managers were young and many of them had changed their job during the follow-up time, which may have caused a variance in burnout scores over time. In future, the rank-order stability of the BBI-9 scores should be investigated in caution by taking into account participants' job changes as well as changes in their nature of work (e.g., autonomy, job control, promotions).

Conclusions

To conclude, there are several reasons why scholars should consider the use of the BBI-9 in their studies on burnout. First, the BBI-9 is a brief measure including the three fundamental aspects of burnout. Second, all items of the BBI-9 are worded negatively and it thus avoids the

criticism faced by the MBI. Third, the three-dimensional structure of the BBI-9 remained the same regardless of the sample differences (group invariance) or measurements over time (time invariance). These factor analytic findings indicate that the possible mean differences in the BBI-9 scores between groups or across time are not caused by the structural instability of the scale. Thus, we consider in line with Salmela-Aro *et al.*[1)] that the BBI-9 is a promising tool to be used both in occupational health care and scientific studies.

Acknowledgement

The study was funded by the Finnish Work Environment Fund (project 110104).

References

1) Salmela-Aro K, Rantanen J, Hyvönen K, Tilleman K, Feldt T (2011) Bergen Burnout Inventory: reliability and validity among Finnish and Estonian managers. Int Arch Occup Environ Health **84**, 635–45.
2) Maslach C, Jackson S, Leiter M (1996) MBI: Maslach Burnout Investory manual, 3rd Ed., Consulting Psychologist Press, Palo Alto.
3) Maslach C, Leiter MP (1997) The truth about burnout. How organizations cause personal stress and what to do about it. Jossey-Bass Publishers, San Francisco.
4) Mäkikangas A, Hätinen M, Kinnunen U, Pekkonen M (2011) Longitudinal factorial invariance of the Maslach Burnout Inventory-General Survey among employees with job-related psychological health problems. Stress Health **27**, 347–52.
5) Maslach C, Schaufeli WB, Leiter MP (2001) Job burnout. Annu Rev Psychol **52**, 397–422.
6) Schaufeli WB, Enzmann D (1998) The burnout companion to study and practice: A critical analysis. Taylor & Francis, London.
7) Shirom A (2003) Job-related burnout. In: Handbook of occupational health psychology, Quick JS, Tetrick LE (Eds.), 245–265, American Psychological Association, Washington, DC.
8) Maslach C, Jackson SE (1981) The measurement of experienced burnout. J Occup Behav **2**, 99–113.
9) Ahola K, Honkonen T, Isometsä E, Kalimo R, Nykyri E, Koskinen S, Aromaa A, Lönnqvist J (2006) Burnout in the general population. Results from the Finnish Health 2000 Study. Soc Psychiatry Psychiatr Epidemiol **41**, 11–7.
10) Enzmann D (2005) Burnout and emotions: an underresearched issue in search of a theory. In: Research companion to organizational health psychology, Antoniou A-S, Cooper CL (Eds.), 495–502, Edgar Elgar Publishing, Cheltenham.
11) Demerouti E, Verbeke W, Bakker A (2005) Exploring the relationship between a multidimensional and multifaceted burnout concept and self-rated performance. J Manage **31**, 186–209.
12) Toppinen-Tanner S (2011) Process of burnout: Structure, antecedents, and consequences. People and Work Research Reports 93, Finnish Institute of Occupational Health, Helsinki.
13) Schaufeli WB, Salanova M (2007) Efficacy or inefficacy, that's the question: burnout and work engagement, and their relationships with efficacy beliefs. Anxiety Stress Coping **20**, 177–96.
14) Le Blanc PM, Hox JJ, Schaufeli WB, Taris TW, Peeters MCW (2007) Take care! The evaluation of a team-based burnout intervention program for oncology care providers. J Appl Psychol **92**, 213–27.
15) Matthiesen S, Dyregrov A (1992) Empirical validation of the Bergen Burnout Indicator. Int J Psychol **27**, 497.
16) Näätänen P, Aro A, Matthiesen S, Salmela-Aro K (2003) Bergen Burnout Indicator 15. Edita, Helsinki.
17) Hyvönen K, Feldt T, Salmela-Aro K, Kinnunen U, Mäkikangas A (2009) Young managers' drive to thrive: a personal work goal approach to burnout and work engagement. J Vocat Behav **75**, 183–96.
18) Hyvönen K, Feldt T, Tolvanen A, Kinnunen U (2010) The role of goal pursuit in the interaction between psychosocial work environment and occupational well-being. J Vocat Behav **76**, 406–18.
19) Mäkikangas A, Feldt T, Kinnunen U, Tolvanen A (2012) Do low burnout and high work engagement always go hand in hand?: investigation of the energy and identification dimensions in longitudinal data. Anxiety Stress Coping **25**, 93–116.
20) Salmela-Aro K, Näätänen P, Nurmi JE (2004) The role of work-related personal projects during burnout interventions: a longitudinal study. Work Stress **18**, 208–30.
21) Siegrist J (1996) Adverse health effects of high-effort/low-reward conditions. J Occup Health Psychol **1**, 27–41.
22) Siegrist J, Starke D, Chandola T, Godin I, Marmot M, Niedhammer I, Peter R (2004) The measurement of effort-reward imbalance at work: European comparisons. Soc Sci Med **58**, 1483–99.
23) Feldt T, Huhtala M, Kinnunen U, Hyvönen K, Mäkikangas A, Sonnentag S (2013) Long-term patterns of effort-reward imbalance and over-commitment: investigating occupational well-being and recovery experiences as outcomes. Work Stress **27**, 64–87.
24) Muthén LK, Muthén BO (1998-2011) Mplus User's guide, 6th Ed., Muthén & Muthén, Los Angeles.
25) Little TD (1997) Mean and covariance structures (MACS) analyses of cross-cultural data: practical and theoretical issues. Multivariate Behav Res **32**, 53–76.
26) Satorra A, Bentler PM (2001) A scaled difference chi-square test statistic for moment structure analysis. Psychometrika **66**, 507–14.

27) Kelloway EK (1998) Using LISREL for Structural Equation Modeling: A researcher's guide. Sage Publications, Thousand Oaks.
28) Steiger JH (1990) Structural model evaluation and modification: an interval estimation approach. Multivariate Behav Res **25**, 173–80.
29) Hu L, Bentler PM (1999) Cutt-off criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling **6**, 1–55.
30) MacCallum RC, Browne MW, Cai L (2006) Testing differences between nested covariance structure models: power analysis and null hypotheses. Psychol Methods **11**, 19–35.
31) Kalimo R, Hakanen J, Toppinen-Tanner S (2006) Maslachin yleinen työuupumuksen arviointimenetelmä *MBI-GS* [The Finnish version of Maslach's Burnout Inventory – General Survey]. Työterveyslaitos, Helsinki.
32) Brandes P, Castro S, James M, Martinez A, Matherly T, Ferris G, Hockwarter W (2008) The interactive effects of job insecurity and organizational cynicism on work effort following a layoff. J Lead Org Stud **14**, 233–47.
33) Shirom A, Melamed S (2006) A comparison of the construct validity of two burnout measures in two groups of professionals. Int J Stress Manag **13**, 176–200.