Translation and validation of the Copenhagen Burnout Inventory amongst Greek doctors

Efthathios Papaefstathiou,1 Andreas Tsounis,2 Maria Malliarou,3 Pavlos Sarafis4

12nd Urology Department, School of Medicine, Aristotle University of Thessaloniki, Greece; 2School of Psychology, Aristotle University of Thessaloniki, Greece; 3Faculty of Nursing, Technological Educational Institute of Larissa, Greece; 4Department of Nursing, School of Health Sciences, Limassol, Cyprus

Abstract

The Copenhagen Burnout Inventory (CBI) is a tool for burnout measurement and includes three different subscales: (i) personal, (ii) work-related and (iii) client-related burnout. The aim of this study was the translation and validation of CBI into the Greek Language. The forward-backward translation method was performed. Initially the questionnaire was distributed to 35 medical students for face validity assessment. Then, 284 residents answered the Greek version of CBI so as for construct validity to be examined. The data analysis was performed by using SPSS and AMOS. Face validity was estimated above 0.8 for the Greek version of CBI. The three factors model of CBI-Greek, with 1 item removed, achieved not so well-defined inner structure in CFI. However, the 16-item model achieved good levels of goodness-of-fit indices (Cmin/df 2.52, RMSEA 0.074, GFI 0.901, CFI 0.938, NFI 0.901, TLI 0.923). The composite reliability values, also varied from 0.842 to 0.852. Cronbach’s alpha index values surpassed 0.844. The 16-item model of CBI-Greek is a valid tool with high internal consistency that can be used in the evaluation of burnout in Greek population. The original 3-items model is a possible alternative.

Introduction

Burnout syndrome refers to a psychological response to chronic work-related condition. It affects both employees’ well-being leading to physical weakness, insomnia, anxiety and depression and the institution’s function leading to increased absenteeism and poor work performance (Gómez-Urquiza et al., 2017; Lahana et al., 2017; Parola et al., 2017). Multiple tools have been constructed for the evaluation of burnout, namely the Maslach Burnout Inventory (MBI), the Oldenburg Burnout Inventory (OBI), the Burnout Measure (BM) and the Copenhagen Burnout Inventory (CBI) (Shirom, 2005; Schaufeli et al. 2009; Halbesleben & Demerouti, 2005).

Amongst them, the most popular and well-used is the MBI. This tool includes three different subscales for the evaluation of burnout: emotional exhaustion, depersonalization and reduced sense of personal accomplishment (Schaufeli & Buunk, 2002; Schaufeli & Taris, 2005). However, although MBI is the most popular measurement tool of burnout it has been the subject of a considerable debate. MBI is highly criticized for not been developed from a sound theoretical model (Platsidou & Daniilidou, 2016; Shirom, 2005). Many researchers indicate that depersonalization is a coping strategy and not a component of burnout syndrome. In addition, many studies revealed that personal accomplishment can be developed independently from the other MBI dimensions (Platsidou & Daniilidou, 2016; Shirom, 2005). For this reason, research on its’ psychometric properties, concerning alternative tested models of the inventory, lead to increased or decreased factors or items compared to the original 22-item set (Worley et al., 2008; Brisson & Bianchi, 2017; Maticorena-Quevedo et al., 2016; Bianchi, 2015; Loera et al., 2014). Moreover, MBI has been a lot criticized as a tool that can only be applied among employees in the human services sector (Kristensen et al., 2005).

As a result to the above criticism Kristensen et al. (2005) developed CBI. The main advantage of CBI is the fact that it can evaluate burnout in a greater variance of occupations (Milfon et al., 2008; Lee et al., 2005). According to their theoretical approach burnout is conceptualized solely as a state of exhaustion
which can be both general (personal burnout) and specific (work and client burnout) (Shirom, 2005; Schaufeli et al., 2009; Kristensen et al., 2005). More specifically, CBI consists of 3 different domains: (i) personal, (ii) occupational and (iii) client-related burnout. The first one refers to physical and psychological fatigue that accumulates in a person during the day. The questions involved are more general and comprehensive from all participants irrespective of occupation. The level of exhaustion and fatigue that derive from work is investigated in the second subscale. The exhaustion that is perceived as a consequence of interpersonal relation with clients is examined by client-related burnout questions. The term client is more general and refers to all types of service recipients such as students, patients, clients etc. This domain of the CBI assesses the exhaustion originating from people centered professions (Kristensen et al., 2005).

Exhaustion is conceptualized as the key component of burnout not only in the case of CBI but also in the development of other tools evaluating burnout. One such tool is OLBI that examines two main dimensions: disengagement and fatigue (physical and cognitive exhaustion) (Demerouti et al., 2003; Demerouti & Bakker, 2008). Another tool is Burnout Measure (BM) of Pines & Aronson, (1988) that evaluates physical, emotional and mental exhaustion. In spite of its' multidimensional structure BM compute a single score for the total scale while research for its' psychometric properties revealed contradictory findings as far the factor solution of the instrument. Similarly, studies conducted for testing OLBI psychometric properties show that the construction of the factors differed from that in the original version of the tool (Baka & Basinska, 2016; Da Silva Schuster & daVeiga Dias, 2018).

In contrast, several studies have supported both the validity and reliability of CBI (Yeh et al., 2007; Berat et al., 2016). Moreover, multiple translations have taken place up until now, including English, Mandarin, French, Japanese, Slovenian and Persian languages while it has also been validated in Brazil, Portugal, Denmark, Australia, Taiwan, Spain, Italy and more (Molinero-Ruiz et al., 2013; Fong et al., 2014; Mahmoudi et al., 2017). Previous findings support the original structure of the tool in Danish, Australian (Winwood & Winefield, 2004), New Zealand (Milfont et al., 2008) and Japan (Odagiri et al., 2004) samples but yet numerous studies do not report its; psychometric properties.

The aim of this study was the translation and validation of CBI for the Greek population in order to become a useful tool in the assessment and comparison of burnout amongst Greek professionals. This purpose is further facilitated by the fact that CBI usage is freely available promoting research in economically deprived countries (Milfont et al., 2008; Lee et al., 2005). The whole validation process examines the clarity and ease of understanding of items, the construct validity and the internal consistency of the questionnaire.

Materials and Methods

Copenhagen Burnout Inventory (CBI)

The CBI consists of 19 items. It evaluates (i) personal related (6 items), (ii) work related (7 items) and (iii) client-related (6 items) burnout. Personal exhaustion refers to both physical and psychological fatigue that accumulates in a person during the day (e.g. “How often are you physically exhausted?”). Occupational exhaustion describes fatigue that is derived from work (e.g. “Do you feel worn out at the end of the working day?”). Client-related exhaustion depicts burnout as a consequence of interpersonal relationship with the clients (e.g. “Does it drain your energy to work with clients?”) (Kristensen et al., 2005).

Translation procedure

Initially a translation of the tool was performed according to the World Health Organization suggested translation guidelines (WHO, 2014). The forward translation of the English version of CBI was performed independently by two translators into Greek. Following that the two versions of CBI were discussed in a meeting. Reconciliation to create one harmonized translation provided the 1st version of CBI-Gr. The expert panel included a doctor, a psychologist and an expert on the particular scientific field and a linguistic teacher. A backward translation of the primary version back into English was performed by an independent translator, whose mother tongue is English and who was not aware of the tool. This process was also followed by a meeting in order to harmonize the 2nd version of CBI with the original English version. Next, a pre-test of the instrument through personal interviews occurred, in order to investigate for unacceptable or offensive words and unclear expressions and choose between alternatives conforms (version 3). The respondents were native Greek speakers (4 male and 4 female) who worked as residents for at least 2 years. The final version was reached by a final translator after taking into consideration the previous problems. The questionnaires were distributed online through Google Forms and demographic data were also recorded.

Data collection

Participants and procedure

Initially 34 medical students answered questions pertaining to the comprehensibility and clarity of the CBI-Gr questions, so as for the face validity to be estimated. Next, the Greek version of CBI was completed by 284 Greek residents working in different hospitals in Greece and abroad. The sample size was based on the 10-20 samples per item suggested for a validation study (Costello & Osborne, 2005). All the participants were native Greek speakers who had completed higher education. The main non-inclusion criterion was not speaking Greek as mother language. Data collection was made online using google platform. The links for the online questionnaire were handed out via medical residents’ emails. From the 345 emails that were sent 284 residents gave their consent to participate to the study and completed the questionnaires (Response rate: 82.32%). Most of the medical residents that received messages were from two hospitals of Thessaloniki, while messages were also sent on junior doctors working in four hospitals of other Greek cities.

Ethical considerations

The study was approved by the scientific councils of the involved hospitals Participants, who met the criteria, after being informed about the aim of the study, gave their consent and completed the above-mentioned questionnaires. Participants’ anonymity was protected and safety of the material was maintained.

Data analysis

The statistical analysis was performed with Statistical Package for Social Sciences (SPSS) version 24 and AMOS 25v. For the purpose of assessing face validity two separate Likert-scales were used to evaluate clarity and comprehension. The former was evaluated with a 5-point scale that varied from 1-not clear at all to 5-
very clear and the latter with a similar scale from 1-totally incomprehensible to 5-easy to understand. Face validity index was the average index value of the above indexes. The results were then converted in values between 0 (totally unclear or incomprehensive) to 1 (clear or understandable). In this study a face validity index above 80% was regarded satisfactory (Polit et al., 2007). Goodness-of-fit indices were used to assess construct validity of the questionnaire. Convergent validity and discriminant validity were also estimated. Absolute fit was measured using Root Mean Square of Error Approximation (RMSEA) and Goodness-of-Fit Index (GFI) with level of acceptance less than 0.08 and more than 0.9 respectively (Browne & Cudeck, 1992; Joreskog & Sorbom, 1984). Incremental fit was calculated using Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) and Normed Fit Index (NFI), all of them expected to be more than 0.9 (Bentler, 1990; Bentler & Bonett, 1980). Last, parsimonious fit was examined by Chi-Squared/ Degree of Freedom (Chi^2/df) which should be less than 5. Size of factor loading, Average Variance Extracted (AVE), with a required value of more than 0.5, and composite reliability (CR), with a required value of 0.6 or more, were used to estimate convergent validity (Hair et al., 2009; Zainudin, 2012). Both AVE and CR were calculated manually. Discriminant validity was also calculated using a comparison between AVE and shared variance (SV) values. Discriminant validity was satisfactory when AVE was higher than SV (correlation < 0.85) (Chin et al., 2018). The reliability of the scale was examined using Cronbach’s alpha coefficient. Values from 0.6 to 0.7 and from 0.7 to 0.9 were indicative of satisfactory and high internal consistency respectively.

### Results

#### Participants’ characteristics

The sample of the study included 284 medical residents. Most of them were males (51.3%), with mean age 31.7 years old. Concerning their family status, the majority were single (71.0%) while 27.1% were married. As far as their specialty, 51.3% of them were trained in the internal medicine sector, 41.6% in the surgical sector and a 7.1% of them attended a specialty in the laboratory sector. Moreover, 39.8% has completed post-graduate studies. Regarding the location of the hospital of medical residency, the majority of them (62.5%) worked in hospital of Thessaloniki, while 37.5% worked in other cities of Greece.

#### Face validity

Pertaining to face validity, universal validity index was 86.42%, while clarity was 85.02% and comprehension 87.81%. Further analysis per subscale and question are shown in Table 1. The aforementioned values indicate sufficient face validity.

#### Construct validity

Evaluating construct validity, the one-factor model which includes 19 questions was not fit, underlining the presence of several constructs. The suggested three-factor model was borderline acceptable in terms of goodness-of-fit indices in the Greek version of CBI. So as for an improvement on the current model to be performed, stepwise removal of items took place. The procedure was based on modification indices, standardized residual covariance and standardized regression as well. The removal of q13 which correlated poorly with occupational burnout did not significantly improved goodness-of-fit indices. Finally, a 3-factor model including 16 items, excluding question 12, 13 from work-related and 17 from client-related burnout, achieved significant model fit (Figure 1). All the goodness-of-fit indices for all models are presented in Table 2.

#### Convergent and discriminant validity

Convergent validity for both models was estimated through calculation of the size of factor loading. Average Variance Extracted (AVE) and Composite Reliability (CR) as seen in Table 3. All the standardized factor loadings surpass 0.5, apart from Q6 and also AVE is above 0.5 for both work-related and client-related burnout. Moreover, composite reliability values are above 0.6 suggesting satisfactory convergent validity (Hair et al., 2009). The calculation of SV, AVE and correlation between constructs underpinned good discriminant validity between client related burnout and both work-related and personal related burnout, but not satis-

---

### Table 1. Clarity, comprehension and face validity total, per question and subscale (%).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
<th>Q8</th>
<th>Q9</th>
<th>Q10</th>
<th>Q11</th>
<th>Q12</th>
<th>Q13</th>
<th>Q14</th>
<th>Q15</th>
<th>Q16</th>
<th>Q17</th>
<th>Q18</th>
<th>Q19</th>
<th>P.</th>
<th>W.</th>
<th>C.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity</td>
<td>85.14</td>
<td>88</td>
<td>85.14</td>
<td>88</td>
<td>88</td>
<td>91.42</td>
<td>82.85</td>
<td>86.28</td>
<td>77.71</td>
<td>85.14</td>
<td>94</td>
<td>81.14</td>
<td>80</td>
<td>87.42</td>
<td>86.28</td>
<td>89.71</td>
<td>89.71</td>
<td>80</td>
<td>79.42</td>
<td>87.61</td>
<td>82.44</td>
<td>85.42</td>
</tr>
<tr>
<td>Comprehension</td>
<td>93.14</td>
<td>82.28</td>
<td>81.14</td>
<td>86.28</td>
<td>96.57</td>
<td>93.14</td>
<td>86.28</td>
<td>83.42</td>
<td>88.57</td>
<td>87.42</td>
<td>89.14</td>
<td>84.57</td>
<td>87.42</td>
<td>87.42</td>
<td>90.85</td>
<td>81.71</td>
<td>83.14</td>
<td>91.14</td>
<td>89.71</td>
<td>86.28</td>
<td>88.76</td>
<td>86.69</td>
</tr>
<tr>
<td>Face validity</td>
<td>89.14</td>
<td>85.14</td>
<td>83.14</td>
<td>87.14</td>
<td>92.28</td>
<td>92.28</td>
<td>84.57</td>
<td>84.85</td>
<td>83.14</td>
<td>86.28</td>
<td>86.57</td>
<td>82.85</td>
<td>83.71</td>
<td>87.42</td>
<td>88.57</td>
<td>85.71</td>
<td>85.14</td>
<td>86.14</td>
<td>82.85</td>
<td>85.14</td>
<td>86.85</td>
<td>88.19</td>
</tr>
</tbody>
</table>

*Personal (P); Q1-Q6: work-related (W): Q8-Q11, client-related (C): Q14-18.

### Table 2. Confirmatory analysis of the Greek version of CBI.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Χ² statistic (df)</th>
<th>p-value</th>
<th>Cmin/df</th>
<th>RMSEA</th>
<th>GFI</th>
<th>CFI</th>
<th>NFI</th>
<th>TLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-factor model</td>
<td>1061.85 (152)</td>
<td>&lt;0.001</td>
<td>6.98</td>
<td>0.146</td>
<td>0.645</td>
<td>0.68</td>
<td>0.648</td>
<td>0.641</td>
</tr>
<tr>
<td>Three-factor model</td>
<td>340.34 (145)</td>
<td>&lt;0.001</td>
<td>2.34</td>
<td>0.069</td>
<td>0.887</td>
<td>0.931</td>
<td>0.887</td>
<td>0.918</td>
</tr>
<tr>
<td>Three-factor model</td>
<td>319.55 (128)</td>
<td>&lt;0.001</td>
<td>2.49</td>
<td>0.073</td>
<td>0.888</td>
<td>0.931</td>
<td>0.891</td>
<td>0.918</td>
</tr>
<tr>
<td>Three-factor model</td>
<td>246.91 (98)</td>
<td>&lt;0.001</td>
<td>2.52</td>
<td>0.074</td>
<td>0.901</td>
<td>0.937</td>
<td>0.901</td>
<td>0.923</td>
</tr>
</tbody>
</table>

*aBased on the proposed construct, 19 items. bBased on the proposed construct, 18 items, questions 13 was removed from the original construct. cBased on the final model, 16 items, questions 12, 13 and 17 were removed from the original construct.*
factory discriminant validity between personal and work-related burnout.

Reliability

In reliability analysis, Cronbach’s alpha exceeds 0.7 for all subscales indicating a high level of internal consistency (Table 4).

Discussion

The forward-backward translation followed WHO’s guidelines. In the process language experts in both Greek and English and content experts were included. A pilot study to investigate for language barriers was held and the additional information and corrections were incorporated in the final model. Reconciliation meetings between researchers were held in each step, in order to evaluate the translation result.

In terms of face validity of the CBI-Gr, all questions were understandable and clear with index values surpassing 80%. Pertaining to the structure of questionnaire, the 3-structure model was fit, as indicated from previous studies as well (Kristensen et al., 2005; Milfont et al., 2008; Yeh et al., 2014).

Construct validity was satisfactory after removal of 1 item and further improved for all goodness-of-fit indices after removal of 3 items. Convergent validity was also adequate. Additionally, discriminant validity was satisfactory, except from personal and work-related burnout. With respect to the items removed, question 13 (Do you have enough energy for family and friends during leisure time?) has also been reported as problematic in terms of validity from other 3 studies and could probably become a matter of further investigation due to its poor correlation with work-related burnout or possible gender bias (Yeh et al., 2007; Berat et al., 2016; Chin et al., 2018). However, the removal of item 13 could

Table 3. SV and AVE of CBI based on final model.

<table>
<thead>
<tr>
<th>Factors</th>
<th>AVE</th>
<th>SV by factor</th>
<th>Cronbach’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Personal</td>
<td>Work-related</td>
</tr>
<tr>
<td>Personal</td>
<td>0.4794</td>
<td>0.8686</td>
<td>0.1806</td>
</tr>
<tr>
<td>Work-related</td>
<td>0.5277</td>
<td>0.8686</td>
<td></td>
</tr>
<tr>
<td>Client-related</td>
<td>0.5396</td>
<td>0.1806</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Convergent validity and reliability of the 16-item CBI-Gr based on the final model.

<table>
<thead>
<tr>
<th>Item</th>
<th>Standardized factor loading</th>
<th>Domain</th>
<th>AVE</th>
<th>CR</th>
<th>Cronbach’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1.</td>
<td>0.736</td>
<td>Personal</td>
<td>0.4794</td>
<td>0.8428</td>
<td>0.845</td>
</tr>
<tr>
<td>Q2.</td>
<td>0.644</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3.</td>
<td>0.606</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4.</td>
<td>0.785</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5.</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6.</td>
<td>0.481</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q7.</td>
<td>0.619</td>
<td>Work-related</td>
<td>0.5277</td>
<td>0.8442</td>
<td>0.844</td>
</tr>
<tr>
<td>Q8.</td>
<td>0.882</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9.</td>
<td>0.571</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q10.</td>
<td>0.847</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q11.</td>
<td>0.659</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q12.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q13.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q14.</td>
<td>0.717</td>
<td>Client-related</td>
<td>0.5396</td>
<td>0.8526</td>
<td>0.845</td>
</tr>
<tr>
<td>Q15.</td>
<td>0.758</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q16.</td>
<td>0.819</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q17.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q18.</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q19.</td>
<td>0.575</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Question 12, 13 and 17 were removed from the original construct. aReliability analysis; Cronbach’s alpha coefficient, overall Cronbach’s alpha 0.901. bDomains were predetermined based on a previous study. cAVE (average variance extracted) was calculated manually. dCR (composite reliability) was calculated manually.
not solely address the validity issues. Further exclusion of item 12 (Do you feel that every working hour is tiring for you?) and 17 (Do you feel that you give more than you get back when you work with clients?) based on standardized residual covariance created our final model with improved fit. More than that, significant correlation and inadequate discriminant validity between personal and work-related burnout was noted. This finding suggests an essential overlap between the two constructs, a fact underlined also from other studies (Chin et al., 2018; Fong et al., 2014). As regards internal consistency, studies from several countries, ours included, consider CBI a stable tool. For the reasons above, we suggest the use of the CBI-Gr in general population. The 16-item model has satisfactory construct validity and could be used in future research. The usage of the 18-item model could also be sufficiently applied. Similar studies, where question 12 was removed, suggest the use of the initial 3-factor model with precaution interpreting that particular item. This proposition could facilitate the comparison among different cultures and linguistic groups (Fong et al., 2014).

Our results encourage the broader use of this questionnaire in the evaluation of burnout and fatigue. On account of its sufficient psychometric properties CBI-Gr could be a valuable tool in assessing burnout without mixing it with coping mechanisms or the syndrome’s consequences (Yeh et al., 2007; Berat et al., 2016). Due to focusing exclusively on exhaustion CBI facilitates the discovery of causative factors of the syndrome and allows elaborating treatment strategies ( Yusoff, 2014).

Limitations of the study

The above study has certain limitations. First, it was a convenience sample in a particular occupational group so the results cannot be generalized. Second the sample size was relatively small. Third, we did not use other research tools in order to estimate convergent and discriminant validity. Finally, we did not test a test retested reliability method to examine the consistency over time.

Conclusions

The CBI-Gr is a valid inventory with good psychometric properties assessed in a group of doctors. Due to its flexibility it is proposed for usage in future studies in Greek population especially on account of its targeting on exhaustion and fatigue. The widespread of CBI-Gr will allow the evaluation of burnout among Greek population. Additionally, the inventory is free-of-charge promoting research and comparison among different populations, since it has already been translated in several languages and validated in several groups ( Kristensen et al., 2005; Milfont et al., 2008; Berat et al., 2016). Lastly, this is the first study evaluating the Greek version of CBI in Greek doctors and suggesting a 16-item model with a well-defined inner structure.

References


