Empirical Article

The psychometric properties of the Burnout Assessment Tool in Norway: A thorough investigation into construct-relevant multidimensionality

Leon T. De Beer,1,2 Marit Christensen,3 Torhild A. Sørengaard,3 Siw T. Innstrand1 and Wilmar B. Schaufeli4,5

1Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway
2WorkWell Research Unit, North-West University, Potchefstroom, South Africa
3Department of Leadership and Organizational Behaviour, BI Norwegian Business School, Trondheim, Norway
4Department of Psychology, Utrecht University, Utrecht, the Netherlands
5Research Unit Occupational & Organizational Psychology and Professional Learning, KU Leuven, Leuven, Belgium


Objective
The World Health Organization recognizes burnout as an occupational issue. Nevertheless, accurately identifying employee burnout remains a challenging task. To complicate matters, current measures of burnout have demonstrated limitations, prompting the development of the Burnout Assessment Tool (BAT). Given these circumstances, conducting an in-depth examination of the BAT’s construct-relevant multidimensionality is crucial.

Method
This study focuses on both the original 23-item BAT and the short 12-item version, using modern factor analytic methods to investigate reliability, validity, and measurement invariance in a representative sample from Norway (n = 493; 49.54% women).

Results
Our findings revealed that the bifactor exploratory structural equation modeling solution (burnout global factor and four specific burnout component factors) best explained the data for both BAT versions. All factors demonstrated adequate omega coefficients, with the global factor showing exceptional strength. Both BAT versions correlated highly with each other and with another burnout measure, suggesting convergent validity. Furthermore, both BAT versions achieved full (strict) measurement invariance based on gender. Finally, our results showed that burnout acts as a mediator in our proposed job demands–resources model as preliminary evidence of predictive validity.

Conclusions
The study validates the Burnout Assessment Tool in the Norwegian context. The study supports the reliability, validity, and unbiased nature of the tool across genders. The findings also reinforce the importance of job demands and resources, along with burnout as a key mediator, in understanding workplace dynamics in accordance with job demands–resources theory.

Key words: Burnout, burnout assessment tool, exploratory structural equation modeling, measurement invariance, job demands, job resources.

Leon T. De Beer. Department of Psychology, Dragvoll Campus, Edvard Bull's veg 1, Trondheim 7491, Norway. Email: leondb@gmail.com

INTRODUCTION
The aftermath of the COVID-19 pandemic has sparked renewed emphasis on employee mental health and well-being (Peters, Dennerlein, Wagner & Sorensen, 2022). In the occupational health psychology domain, no topic is perhaps as popular as burnout, which has been the subject of research for almost half a century. The World Health Organization (WHO) acknowledged burnout as a syndrome stemming from chronic workplace stress that is difficult to manage (WHO, 2019). However, burnout has not been classified as a medical condition but rather as an occupational phenomenon (WHO, 2019). One possible explanation for the difficulty in classifying a case of burnout includes the inherent challenges in its diagnosis. While some countries, like the Netherlands, have established national guidelines for assessing burnout, there are no standardized, globally accepted diagnostic criteria for making a formal diagnosis (Nadon, De Beer & Morin, 2022; Parker & Tavella, 2022). This absence of uniform criteria means that occupational health practitioners often depend on self-report surveys to identify employees at risk of burnout. These surveys provide only a risk level, not a categorical diagnosis, but remain a crucial initial step in identifying employees in distress. If the survey results indicate potential burnout risk, practitioners can then refer employees for further evaluation and support, which may include employee assistance programs or other relevant support services. This approach highlights the critical role that self-report measures can play in identifying organizational burnout risk. Therefore, it is essential that these measures show robust psychometric properties in the context in which they are used by the appropriate professionals.

However, over the last few decades burnout measures have been shown to have some limitations, including differences in the conceptualization of burnout being measured (exhaustion-only or multidimensional [e.g., including cynicism, professional efficacy]; see Guseva-Canu et al., 2021 and Schaufeli, 2021), the prescribed factor structure, which remains unclear (three-factor vs. two-factor; e.g., Worley, Vassar, Wheeler, & Barnes, 2008), the apparent divergent role of professional efficacy (e.g., De Beer &
Bianchi, 2019), and the inability of these measures to provide a global burnout score as is ideally required when measuring a syndrome (Schaufeli, Desart & De Witte, 2020).

THE BURNOUT ASSESSMENT TOOL: DEVELOPMENT AND CONVERGENT VALIDITY

Due to these afore-mentioned limitations, Schaufeli, Desart, and De Witte (2020) endeavored to create a new instrument, the Burnout Assessment Tool (BAT). The development of the BAT is unique in that, unlike the development of other burnout measures, both inductive and deductive approaches were used to decide on the final list of items included in the survey. The specific inductive approach employed by Schaufeli and colleagues is uncommon, because the Dutch authorities acknowledge burnout as an occupational disease, and therefore experts who have experience working with (and categorizing) these patients could be interviewed. As for the deductive phase, a list of 357 items and 66 dimensions for burnout were considered. The initial list was constructed by examining 12 existing burnout questionnaires by means of a literature review (see Schaufeli, Desart & De Witte, 2020, for the exact procedure). After analyzing the list, the authors concluded that (1) exhaustion is the core of burnout, (2) all multidimensional scales included both exhaustion and mental distance, (3) positively phrased items are the exception, and (4) Likert scales are used with between four and seven anchors. Considering these conclusions and expert discussions between the authors, an initial pool of 33 items before formal factor analyses were chosen.

Subsequently, after the factor analyses of the development study, the BAT-assessed burnout syndrome showed 23 items measuring four components at its core; three of these components relate to the inability to invest energy: exhaustion, cognitive impairment, and emotional impairment, which represent extreme tiredness, loss of cognitive control, and loss of emotional control, respectively. The fourth component, mental distance, represents an unwillingness to invest energy (Schaufeli, Desart & De Witte, 2020; Schaufeli & Taris, 2005). This culminated in an updated definition of burnout as “a work-related state of exhaustion that occurs among employees, characterised by extreme tiredness, reduced ability to regulate cognitive and emotional processes, and mental distancing” (Schaufeli, Desart & De Witte, 2020, p. 4). Notably, the BAT-assessed burnout definition does not include a professional (in)efficacy component, in line with past research questioning its role (e.g., De Beer & Bianchi, 2019). Consequently, the four components of the BAT (three energy components and a withdrawal component) align with the theoretical conceptualization of burnout containing components of both exhaustion and mental distance (Schaufeli, 2021; Schaufeli & Taris, 2005).

However, the BAT must demonstrate that it is on a par with other burnout measures by means of convergent validity. Convergent validity considers whether different methods of measuring the same concept produce similar results (Abma, Rovers & van der Wees, 2016), and past research on the BAT in other contexts has shown its convergent validity with other well-known burnout measures: the Oldenburg Burnout Inventory (OLBI) and the Maslach Burnout Inventory (e.g., De Beer, Schaufeli & Bakker, 2022; De Beer, Schaufeli & De Witte, 2022; Schaufeli, Desart & De Witte, 2020).

CROSS-CULTURAL VALIDITY AND NORWEGIAN WORKING LIFE

The 23-item version of the BAT (BAT-23) has shown strong evidence for cross-cultural validity with studies of measurement invariance from South Africa (De Beer, Schaufeli & Bakker, 2022; De Beer, Schaufeli & De Witte, 2022), Austria, Belgium, Germany, Ireland, Finland, the Netherlands, and Japan (De Beer et al., 2020). Moreover, the 12-item short version (BAT-12) has also shown the requisite validity and reliability (Hadžibajramović, Schaufeli & De Witte, 2022; Mazzetti et al., 2022; Oprea, Iliescu & De Witte, 2021) and evidence of equivalence in other contexts (e.g., De Beer, Schaufeli & Bakker, 2022; De Beer, Schaufeli & De Witte, 2022; Sinval, Vazquez, Hutz, Schaufeli & Silva, 2022).

Studies have yet investigated the psychometric properties and the measurement invariance of the BAT in Norway, and it is important to do so. Norway is an egalitarian society, with great equality between the different social groups. Income inequality is smaller than in most other countries, the education system is generally open to all, and everyone has equal rights to medical care and hospitals.

Norwegian working life is characterized by the Nordic model (Gustavsen, 2011), including a high degree of participation and a tripartite collaboration between the parties (workers, unions, and employers) in working life. Norway is one of the countries with the highest employment rates among both women and men. Few countries have higher female labor force participation than Norway. Consequently, demonstrating measurement invariance of the BAT across genders is crucial to ensure its equitable application and to underpin valid comparisons of burnout levels between genders (see Meuleman et al., 2023). In fact, adequate measurement invariance based on gender has previously been found in countries such as Brazil and Portugal (Sinval, Vazquez, Hutz, Schaufeli & Silva, 2022), Croatia (Tomas et al., 2023), and South Africa (De Beer, Schaufeli & Bakker, 2022; De Beer, Schaufeli & De Witte, 2022).

Regarding mental health, Norway does well compared with the EU average. Norway scores among the countries with the lowest prevalence of self-reported anxiety, and there is a lower proportion of workers in Norway than in the EU who are mentally exhausted after work. In terms of mental well-being, Norway is on a par with the EU average. The proportion reporting that their health is at risk because of their work is somewhat lower in Norway than in the EU (STAMI, 2023). While Norwegian employees generally report a better work-life balance than the EU average (STAMI, 2023), many employees are still in distress due to work, and thus it remains an important issue to research.

THE JOB DEMANDS–RESOURCES APPROACH TO BURNOUT

Arguably one of the more popular models in the work and organizational psychology domain to explain how the work environment may affect employee health is the job demands–
resources (JD-R) model (Bakker & Demerouti, 2007; Bakker, Demerouti & Sanz-Vergel, 2023). This model has been applied to the situation in Norway (Kaiser, Patras, Adolfsen, Richardsen & Martinussen, 2020). In the JD-R model, burnout is part of what is called a "health impairment process" that explains the development of burnout mainly as an imbalance between job demands and job resources, creating stress that becomes unmanageable, leading to burnout and eventual undesired outcomes such as sleep disturbances (Sørensgaard & Saksvik-Lehouillier, 2022), turnover (Sobstad, Pallesen, Bjorvatn, Costa & Hystad, 2021), and psychological health problems (Burke & Mikkelson, 2006). Moreover, burnout has also been shown to negatively impact both job satisfaction (e.g., Kaiser, Richardsen & Martinussen, 2021) and life satisfaction (e.g., Hombrados-Mendieta & Cosano-Rivas, 2013). Therefore, burnout has been found to act as a mediator between job demands, job resources, and expected outcomes (e.g., Hakanen, Bakker & Schaufeli, 2006; Hakanen, Schaufeli & Ahola, 2008; Schaufeli & Bakker, 2004) – and this has also been shown when measured with the BAT (see De Beer, Schaufeli & Bakker, 2022; De Beer, Schaufeli & De Witte, 2022). We investigate these variables in a mediation model as part of an initial investigation into predictive validity – see Fig. 1 for a conceptual model.

Specifically, we use three job demands that have been shown to be important in the Norwegian context: work-home conflict (Innstrand, 2022; Innstrand, Langballe, Espnes, Falkum & Aasland, 2008; Jensen, 2016), workload (Svedahl et al., 2019), and emotional load (Langballe, Innstrand, Hagtvet, Falkum & Gjerlov Aasland, 2009). Contrastingly, we focus on role clarity (Kaiser, Richardsen & Martinussen, 2021) and supervisor support (Martinussen, Richardsen & Burke, 2007) as job resources. To our knowledge, this is the first study to investigate these factors’ relationship with BAT-assessed burnout.

RESEARCH OBJECTIVE AND HYPOTHESES

Recognizing the critical importance of culturally and contextually relevant diagnostic tools, it becomes imperative to assess the applicability of BAT beyond the contexts in which it has initially been validated. The distinct work-life balance and social systems in Norway, particularly its advanced egalitarian values, may influence the manifestation and measurement of burnout in ways that may not have been captured in previous studies. With the high participation of both genders in the workforce and the specific job demands and resources characteristic of the Norwegian workplace, this research seeks to fill that gap. Therefore, this study aims to investigate the validity, reliability, and measurement invariance of the original BAT-23 and shortened BAT-12 in Norway with a representative sample. The mediating role of burnout in the health impairment process is also considered in the study.

Hypothesis 1: The BAT can be modeled as a total score with four specific component factors.

Hypothesis 2: The BAT shows convergent validity with the Maslach Burnout Inventory – General Survey (MBI-GS).

Hypothesis 3: The BAT shows measurement invariance across gender groups.

Hypothesis 4: Burnout acts, in accordance with the JD-R model, as a mediator between job demands (work-home conflict, workload, emotional demands), job resources (supervisor support, role clarity), and outcome variables (job satisfaction, life satisfaction, and turnover intention).

MATERIALS AND METHODS

Participants and procedure

The data research company Bilendi was used to recruit a representative sample of approximately 500 Norwegian workers according to age and gender to complete a survey on the QuestionPro platform. Therefore, the researchers had no direct contact with the participants. Given that we have access to the anonymous data but not the participants, it is not possible to identify any specific individuals, as both Bilendi and QuestionPro adhere to GDPR regulations. Once participants entered the QuestionPro survey, they had to read and accept the informed consent letter explaining the purpose of the study and confirm their voluntary participation before they could continue. The survey contained two attention check questions, which instructed participants to select a specific answer on the scale to demonstrate that they were still paying attention. In the end, there were 530 complete responses, but 37 participants failed the attention checks. As these data could not be fully trusted, they were excluded, resulting in a final sample of 493 employees (n = 493). However, due to the removal of these participants, the data had to be slightly weighted by age and gender to maintain the representativeness of these categories. The average participant was 45.55 years old (SD = 11.54 years) and had, on average, 2.42 children. We provide a further weighted breakdown of the available participants’ characteristics in Table 1.

Furthermore, using the pooled cutoff criteria for the BAT-23, provided by Schaufeli, De Witte, Hakanen, Kaltiainen, and Kok (2023), the total burnout score of this sample indicates that 12.9% (n = 63) of the sample was at high risk of becoming burned out. However, it must be emphasized that in the continued absence of clinical diagnostic criteria to identify burnout cases, this estimate should be taken not at face value but as a guideline to identify employees who may need further assistance.

Measures

Burnout was assessed with the full 23-item version of the BAT (23; Schaufeli, Desart & De Witte, 2020) and with the shortened 12-item version (BAT-12; Hadžibajramović, Schaufeli & De Witte, 2022). Precisely, the following four underlying aspects were measured: exhaustion, by eight items (e.g., “At work, I feel mentally exhausted”), mental distance, by five items (e.g., “I feel indifferent about my job”), cognitive impairment, by five items (e.g., “When I’m working, I have trouble concentrating”), and emotional impairment, by five items (e.g., “At work, I feel unable to control my emotions”). The BAT-12 comprises three items for each of the four components, and the example items from the
Finally, we also measured the three outcome variables: *job satisfaction* (“Overall, I am satisfied with my job”); a seven-point scale ranging from *Extremely dissatisfied to Extremely satisfied*, *life satisfaction* (“As a whole, I am satisfied with my life”; a seven-point scale ranging from *Extremely dissatisfied to Extremely satisfied*; Fisher, Matthews & Gibbons, 2016), and *turnover intention* (“I am actively looking for other jobs”; a five-point scale ranging from *Strongly disagree to Strongly agree*; Sjöberg & Sverke, 2000).

### Analyses

Mplus 8.9 (Muthén & Muthén, 2023) was used to model the data. Specifically, we used latent variables within a confirmatory factor analysis (CFA) framework for the measurement model with the mean- and variance-adjusted weighted least squares (WLSMV) estimation procedure. First, we systematically estimated six sequential model specifications for each version of the BAT: (1) one-factor CFA, (2) four-factor CFA, (3) second-order factor CFA, (4) bifactor CFA (BCFA), (5) four-factor exploratory structural equation modeling (ESEM), and (6) bifactor exploratory structural equation modeling (BESEM) to ascertain the most appropriate model (Hoyle, 2023). The BESEM model differed from the four-factor ESEM model by adding a global factor to the model and specifying all factors as orthogonal. To assess the adequacy of these models, we employed the comparative fit index (CFI) and the Tucker-Lewis index (TLI), where values of at least 0.90 are indicative of acceptable fit and values exceeding 0.95 are considered to reflect excellent fit. Moreover, for the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR), values up to 0.08 are recognized as denoting acceptable fit (see Kline, 2011; Van de Schoot, Lugtig & Hox, 2012; Wang & Wang, 2020).

Second, based on the most appropriate model for the data, we conducted tests of measurement invariance between genders on both versions of the BAT, including configurual (factor structure), metric (factor loadings), strong (thresholds), and strict invariance (uniquenesses) (Morin, 2023; Patrick & Bornstein, 2016). Strong invariance indicates that the latent mean scores can be fairly compared between groups, whereas strict invariance (full measurement invariance) indicates that observed scores can also be compared between groups. Regarding the evaluation of the tests for measurement invariance, a decline in CFI and TLI values of −0.01 or greater, or an increase in RMSEA of +0.015 or more when juxtaposing a model with the previous model in the sequence, is indicative of a failure to uphold that level of invariance (Chen, 2007; Cheung & Rensvold, 2002). Therefore, if the cutoff criterion is violated, the level of invariance is considered at the previous acceptable level.

To consider the reliability of the scales, we calculated McDonald’s omega coefficients (McDonald, 1970). This approach is especially apt when estimating (BE)SEM models (Morin, 2023). We also tested correlations for convergent validity between the BAT-23, BAT-12, and MBI-GS scores. A correlation approaching one suggests that the constructs are converging (Brown, 2015).

Furthermore, for parsimony, we also tested a structural model (mediation model) based on the observed scores of the variables. From the analyses, we specifically considered the statistical significance (p < 0.05) and direction of the standardized beta coefficients. To test for burnout’s mediating role in the model, we also bootstrapped the model parameters 10,000 times to generate a range of lower and upper 95% confidence intervals for the indirect effects. If a confidence interval did not contain zero (change sign), it indicates that the indirect effect can be seen as meaningful (Hayes, 2022). This implies that the variable is functioning as a mediator in the model.

## RESULTS

### Model fit, correlations, and omega coefficients (composite reliability)

As shown in Table 2, our systematic investigation of the factor structure of both versions of the BAT revealed that a

<table>
<thead>
<tr>
<th>Category</th>
<th>Sub-category</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>247</td>
<td>50.06</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>244</td>
<td>49.54</td>
</tr>
<tr>
<td></td>
<td>Non-binary</td>
<td>1</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>1</td>
<td>0.20</td>
</tr>
<tr>
<td>Education</td>
<td>Primary school</td>
<td>29</td>
<td>5.93</td>
</tr>
<tr>
<td></td>
<td>Secondary school</td>
<td>154</td>
<td>31.17</td>
</tr>
<tr>
<td></td>
<td>3-year higher education</td>
<td>210</td>
<td>42.52</td>
</tr>
<tr>
<td></td>
<td>4-year higher education or more</td>
<td>100</td>
<td>20.38</td>
</tr>
<tr>
<td>Sector</td>
<td>Health care</td>
<td>124</td>
<td>25.23</td>
</tr>
<tr>
<td></td>
<td>Trade, business, commerce</td>
<td>52</td>
<td>10.64</td>
</tr>
<tr>
<td></td>
<td>Industry</td>
<td>31</td>
<td>6.30</td>
</tr>
<tr>
<td></td>
<td>Construction</td>
<td>38</td>
<td>7.78</td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td>59</td>
<td>11.91</td>
</tr>
<tr>
<td></td>
<td>Office administration</td>
<td>55</td>
<td>11.14</td>
</tr>
<tr>
<td></td>
<td>Professional services</td>
<td>38</td>
<td>7.72</td>
</tr>
<tr>
<td></td>
<td>Information technology and media</td>
<td>42</td>
<td>8.46</td>
</tr>
<tr>
<td></td>
<td>Oil and gas</td>
<td>24</td>
<td>4.77</td>
</tr>
<tr>
<td></td>
<td>Finance and insurance</td>
<td>22</td>
<td>4.44</td>
</tr>
<tr>
<td></td>
<td>Missing values</td>
<td>8</td>
<td>1.60</td>
</tr>
<tr>
<td>Employment</td>
<td>Permanent employee</td>
<td>451</td>
<td>91.26</td>
</tr>
<tr>
<td></td>
<td>Temporary employee</td>
<td>32</td>
<td>6.58</td>
</tr>
<tr>
<td></td>
<td>Self-employed</td>
<td>10</td>
<td>2.16</td>
</tr>
<tr>
<td></td>
<td>Full-time equivalent</td>
<td>387</td>
<td>78.28</td>
</tr>
<tr>
<td></td>
<td>% 51–100%</td>
<td>78</td>
<td>15.88</td>
</tr>
<tr>
<td></td>
<td>1–50%</td>
<td>28</td>
<td>5.70</td>
</tr>
<tr>
<td>Remote working</td>
<td>Never, or almost never</td>
<td>292</td>
<td>59.22</td>
</tr>
<tr>
<td></td>
<td>Sometimes</td>
<td>123</td>
<td>24.91</td>
</tr>
<tr>
<td></td>
<td>Often</td>
<td>43</td>
<td>8.76</td>
</tr>
<tr>
<td></td>
<td>Always, or almost always</td>
<td>35</td>
<td>7.11</td>
</tr>
</tbody>
</table>

Note: Weighted counts and percentages. Count rounded to an appropriate integer. Percentages are unchanged.
unidimensional factor structure was not a good fit. Furthermore, even though some other models showed acceptable fit statistics, the bifactor ESEM models best fitted the data, supporting Hypothesis 1. The online supplementary material provides all models’ accompanying factor loadings and uniquenesses (Tables S1 and S2).

However, our selection of the bifactor ESEM model for both versions of the BAT was based not solely on the fit statistics but also on other considerations such as reductions in correlations and the omega coefficients. Specifically, Table 3 shows the size of the correlations between the factors reduced from the CFA model to the ESEM model in both BAT versions (ΔCF $= 0.155$; ΔRM $= 0.119$), providing additional evidence that an underlying global factor is present (Morin, 2023). Moreover, all correlations were statistically significant in the expected directions and showed large effect sizes, except for the correlation between exhaustion (EX) and emotional impairment (EI) in the BAT-12 model, which was a borderline case ($r = 0.494$; medium effect). All omega coefficients were also above 0.700.

The results of the BESEM models revealed a strongly defined and reliable global factor for both versions of the BAT ($\omega_{12} = 0.966$; $\omega_{23} = 0.945$), consistent with the presence of a strong common core to all BAT items. Beyond this global factor, the results also revealed that meaningful specificity remained at the level of the EX ($\omega_{23} = 0.836$; $\omega_{12} = 0.744$), CI ($\omega_{12} = 0.822$; $\omega_{23} = 0.776$), and EI ($\omega_{12} = 0.743$; $\omega_{23} = 0.712$) factors. However, the MD specific factor seemed to retain less specificity in the BAT-12 ($\omega_{23} = 0.455$), where these items primarily seem to reflect participants’ global levels of burnout. Nonetheless, it is crucial to emphasize that this specific factor in the BAT-12 still maintains meaningful specificity by accounting for variance not accounted for in its global burnout factor. Indeed, the specific factor retained more meaningful levels of specificity in the BAT-23 ($\omega_{23} = 0.646$).

### Table 2. Fit statistics for the model specification investigation of the BAT-23 and BAT-12

<table>
<thead>
<tr>
<th>Model (BAT-23)</th>
<th>$\chi^2$</th>
<th>df</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA [90% CI]</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAT-23 One-factor</td>
<td>2432.89</td>
<td>230</td>
<td>0.836</td>
<td>0.819</td>
<td>0.139 [0.134–0.144]</td>
<td>0.089</td>
</tr>
<tr>
<td>BAT-23 Four-factor</td>
<td>1120.21</td>
<td>224</td>
<td>0.933</td>
<td>0.924</td>
<td>0.090 [0.085–0.095]</td>
<td>0.055</td>
</tr>
<tr>
<td>BAT-23 Second-order factor</td>
<td>1082.85</td>
<td>226</td>
<td>0.936</td>
<td>0.928</td>
<td>0.088 [0.082–0.093]</td>
<td>0.055</td>
</tr>
<tr>
<td>BAT-23 Bifactor CFA</td>
<td>666.27</td>
<td>211</td>
<td>0.966</td>
<td>0.959</td>
<td>0.066 [0.060–0.072]</td>
<td>0.041</td>
</tr>
<tr>
<td>BAT-23 ESEM</td>
<td>519.15</td>
<td>167</td>
<td>0.974</td>
<td>0.960</td>
<td>0.065 [0.059–0.072]</td>
<td>0.026</td>
</tr>
<tr>
<td>BAT-23 Bifactor</td>
<td>452.86</td>
<td>148</td>
<td>0.977</td>
<td>0.961</td>
<td>0.065 [0.058–0.071]</td>
<td>0.023</td>
</tr>
</tbody>
</table>

Note: $\chi^2$ = chi-square; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. Graphical representation of the models is provided in Figs. S1–S6.

### Table 3. Descriptive statistics, omega coefficients, and CFA/ESEM correlation matrix

<table>
<thead>
<tr>
<th>Factors</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAT-23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. EX</td>
<td>2.69</td>
<td>0.73</td>
<td>(0.916/0.913)</td>
<td>0.517</td>
<td>0.519</td>
<td>0.554</td>
</tr>
<tr>
<td>2. MD</td>
<td>2.24</td>
<td>0.75</td>
<td>0.744</td>
<td>(0.844/0.819)</td>
<td>0.594</td>
<td>0.522</td>
</tr>
<tr>
<td>3. CI</td>
<td>2.13</td>
<td>0.72</td>
<td>0.676</td>
<td>0.705</td>
<td>(0.931/0.926)</td>
<td>0.607</td>
</tr>
<tr>
<td>4. EI</td>
<td>1.88</td>
<td>0.66</td>
<td>0.700</td>
<td>0.689</td>
<td>0.726</td>
<td>(0.901/0.872)</td>
</tr>
<tr>
<td>BAT-12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. EX</td>
<td>2.62</td>
<td>0.82</td>
<td>(0.858/0.846)</td>
<td>0.656</td>
<td>0.517</td>
<td>0.494</td>
</tr>
<tr>
<td>2. MD</td>
<td>2.18</td>
<td>0.78</td>
<td>0.760</td>
<td>(0.744/0.716)</td>
<td>0.631</td>
<td>0.600</td>
</tr>
<tr>
<td>3. CI</td>
<td>2.12</td>
<td>0.73</td>
<td>0.631</td>
<td>0.731</td>
<td>(0.884/0.881)</td>
<td>0.584</td>
</tr>
<tr>
<td>4. EI</td>
<td>1.78</td>
<td>0.69</td>
<td>0.643</td>
<td>0.730</td>
<td>0.696</td>
<td>(0.865/0.833)</td>
</tr>
</tbody>
</table>

Note: Four-factor CFA correlations below the diagonal; ESEM correlations above the diagonal; omega coefficients on the diagonal in brackets (CFA/ESEM); M = observed weighted mean; SD = observed weighted standard deviation of the mean; all correlations $p < 0.001$; EX = exhaustion; MD = mental distance; CI = cognitive impairment; EI = emotional impairment.
Hypothesis 3. Furthermore, the mean results showed neither
fi
corresponding observed scores. These
fi
BAT-23 and BAT-12 can be compared between sexes
statistics for the direct and indirect paths.
Table 5 provides the standardized beta coef
fi
cients. However, in both
fi
the BAT-23 and BAT-12, men scored lower on the specific
ex
comp
fi
ponent (M23 = –0.576, p < 0.001; M12 = –0.540, p = 0.003) than women, but also on the CI component in the
fi
BAT-12 (M12 = –0.302, p = 0.026).

Direct and indirect path results
Table 5 provides the standardized beta coefficients and related
fi
statistics for the direct and indirect paths.
Specifically, the results showed that work-home conflict
(β23 = 0.523, SE = 0.038, p < 0.001; β12 = 0.486, SE = 0.039, p < 0.001) and emotional demands (β23 = 0.175, SE = 0.043, p < 0.001; β12 = 0.190, SE = 0.046, p < 0.001), but not workload (β23 = 0.894; β12 = 0.708), had positive paths to burnout for both versions of the BAT. In terms of job resources, supervisor support had a negative path to burnout (β23 = –0.110, SE = 0.046, p < 0.001; β12 = –0.123, SE = 0.047, p < 0.001); role clarity had a negative path to burnout only in the BAT-12 (β12 = –0.092, SE = 0.041, p = 0.023) but not in the BAT-23 (p23 = 0.094).
Interestingly, none of the job demands or job resources had significant direct paths to the outcome variable turnover intention (p > 0.05) except supervisor support, which had a negative relationship (β23 = –0.209, SE = 0.048, p < 0.001; β12 = –0.201, SE = 0.047, p < 0.001). The situation was similar for the job satisfaction outcome, for which only supervisor support to turnover intention showed a positive path (β23 = 0.257, SE = 0.049, p < 0.001; β12 = 0.252, SE = 0.050, p < 0.001). Moreover, for life satisfaction, only work-home conflict showed a significant, negative, relationship (β23 = –0.143, SE = 0.071, p = 0.043; β12 = –0.188, SE = 0.069, p = 0.007).

Burnout revealed a significant positive path to turnover intention (β23 = 0.335, SE = 0.056, p < 0.001; β12 = 0.336, SE = 0.052, p < 0.001) and significant negative paths to job satisfaction (β23 = –0.407, SE = 0.065, p < 0.001; β12 = –0.396, SE = 0.066, p < 0.001) and life satisfaction (β23 = –0.277, SE = 0.073, p < 0.001; β12 = –0.205, SE = 0.074, p < 0.001) in both versions of the BAT. All in all, among the direct paths in the model, the results were relatively consistent between the BAT-23 and BAT-12 except for the path of role clarity to burnout in the BAT-23 relative to the BAT-12, described earlier.

The indirect effects showed a similar pattern in which the direct paths were significant. Therefore, Hypothesis 4 was only partially supported, even though most indirect effects were meaningful. For example, work-home conflict and emotional demands indirectly affected turnover through burnout in both versions of the BAT.

An online application was created for Norwegian employees to consider their own burnout risk results, for entertainment purposes, based on the 23-item BAT data from this study to benchmark against. The application can be accessed at the following location: https://theburnout.app/?mod=no.

DISCUSSION
This study investigated the construct-relevant multidimensionality and measurement invariance of the BAT-23 and BAT-12, an updated tool to measure burnout. Results broadly supported that the scales have robust properties: validity and reliability.

First, Hypothesis 1 was supported as it was established that the best representation of the BAT-related data was the BESEM latent variable representation. That is a model that included a strong global burnout factor, with four specific components of the BAT. This is in line with other recent studies on the BAT, which have

© 2023 The Authors. Scandinavian Journal of Psychology published by Scandinavian Psychological Associations and John Wiley & Sons Ltd.
found a bifactor representation to be most suitable in other contexts (e.g., Bąsinska, Gruszczyńska & Schaufeli, 2023; De Beer, Schaufeli & Bakker, 2022; De Beer, Schaufeli & De Witte, 2022). Furthermore, the evidence shows that the BAT can be modeled as a total score, with its subcomponent scores, solving one of the main concerns against other burnout measures purportedly measuring a syndrome.

Furthermore, the correlations from the convergent validity assessment showed that both versions of the BAT are highly correlated with one another – showing convergent validity ($R^2 = 95.84\%$) and supporting Hypothesis 2. This is important, as it shows that the shorter version of the BAT reflects an excellent representation of the scores one would achieve on the full 23-item version. Similarly, but less strongly, both versions of the BAT correlated highly with the MBI-GS, said to be the “gold standard” of burnout measurement (Schaufeli, Desart & De Witte, 2020).

Regarding equivalence, both versions of the BAT achieved full (strict) measurement invariance as pertaining to gender – supporting Hypothesis 3. This means that not only the latent scores but also the observed scores of BAT-assessed burnout can be compared between men and women. This aligns with other research on the BAT, which has shown it to be invariant across countries (De Beer et al., 2020) and in specific contexts (De Beer, Schaufeli & De Witte, 2022). Specifically, the results showed no statistically significant difference in the mean level of global burnout between men and women in this representative sample. This contrasts with other studies, which have found levels of global burnout to be higher among women (e.g., De Beer, Schaufeli & De Witte, 2022). One potential explanation for this result within the Norwegian context could be the more egalitarian nature of working life, so similar scores are apparent. Nevertheless, the higher level of specific exhaustion found among

![Image of the page containing the table and text]
women using both the BAT-23 and the BAT-12 align with previous studies conducted in Norway using the Oldenburg Burnout Inventory (OLBI) as a burnout measure (Innstrand, Langballe, Falkum & Aasland, 2011). Innstrand and colleagues suggested that the gender differences might vary across occupational groups. Unfortunately, we could not test gender differences across occupational groups in the present study, due to under-representation.

Finally, in the structural model, both versions of the BAT performed almost identically in predicting outcomes, in line with the JD-R model. Of the direct paths, it is important to underline the strong effect of work-home conflict in contributing to burnout. Although this aligns with previous findings from Norway (Innstrand, Langballe, Espnes, Falkum & Aasland, 2008), it underlines the need to respect work-home balance even in the Nordic work model (Gustavsen, 2011). Specifically, burnout mediated between work-home conflict and emotional load and the outcome variables, and between job resources and job satisfaction and life satisfaction, supporting Hypothesis 4. However, an interesting finding was that workload did not have any significant direct effects on burnout in the presence of work-home conflict and emotional load. Our cross-sectional result contradicts stronger findings from a meta-analytic study that found a small effect of workload over time (Guthier, Dormann & Voelkle, 2020). This could be explained by fair workloads in Norway, as the meta-analytic effect was already small — or it could be an artifact in this sample. Be that as it may, work-home conflict and emotional load did show effects, and these should be managed in the workplace. Furthermore, the importance of supervisor support as a job resource is also underlined in this study, as it has been in a meta-analysis (Aronsson et al., 2017).

In general, all the results supported the role of BAT-assessed burnout in the health impairment process of the JD-R model (Bakker, Demerouti & Sanz-Vergel, 2023). Specifically, job demands in the form of work-home conflict and emotional load contributed to burnout and the outcomes of this study, whereas supervisor support as a job resource minimized the mediating impact of burnout on the outcomes. Therefore, leaders of organizations in Norway should be cognizant of these demands and resource to promote optimal outcomes.

LIMITATIONS
This study is subject to limitations that merit attention. While the demographic composition of the sample approximated the general Norwegian population in terms of age group and gender, it remained cross-sectional in nature. Longitudinal research designs are recommended for future studies to establish causality among the variables of interest. Moreover, due to the plurality of healthcare workers in this sample, it would be beneficial for future research to expand sample sizes and aim to include occupational group as an additional dimension of representativeness of the Norwegian workforce. This approach would also facilitate in-depth explorations of intra- and inter-occupational group dynamics by using multigroup analyses and similar techniques, which may yield insightful dynamics about burnout. Second, this study did not include any self-reported depression-related variables that could have been used to contribute to the burnout-depression debate (e.g., Nadon, De Beer & Morin, 2022). Lastly, our data did not contain objective data such as actual performance or productivity, which would have been an interesting feature. Therefore, future studies should endeavor to obtain and include performance, sickness absence, and health-related data to investigate the finer nuances of BAT-assessed burnout in individual and organizational outcomes. These efforts could include investigating at what levels of burnout severity or score the BAT may be useful to predict depression, anxiety, or other health-related outcomes that have binary, diagnostic classifications.

CONCLUSIONS
This study produced robust evidence for the validity, reliability, and gender equivalence of both the BAT-23 and BAT-12 within Norway. Therefore, the BAT-23 and BAT-12 can be used with confidence to ascertain organizational employee burnout risk. The results showed that the BAT-23 captures nuances beyond the global burnout score more thoroughly than the shorter BAT-12. Specifically, with the BAT-12, mental distance (withdrawal) is reflected more by the global burnout factor even though the specific mental distance does maintain some meaningful specificity beyond the global factor. However, the primary advantage of the BAT is its deference to a total score. Our results also supported BAT-assessed burnout’s role as a mediator in the health impairment process of a JD-R model.

ACKNOWLEDGEMENTS
None.

CONFLICTS OF INTEREST
None declared.

ETHICS STATEMENT
Respondents gave informed consent by accepting the voluntary nature of participation before the online survey commenced and were free to cease participation at any time. Due to online anonymity and a limited set of background information, it was not possible for the researchers to identify who was responding to the survey. The project therefore followed the requirements of the Norwegian Centre for Research Data (NSD).

DATA洪AILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

ENDNOTES
1 The MBI-GS is copyrighted, and therefore example items are not shared here.
2 Delta mean (change) in the correlations from the CFA model to the ESEM model.
3 Due to the division of item-level true score variance between two sets of factors (G and S) in the bifactor ESEM solutions, it is common for the S factors to have weaker definition compared with the CFA or ESEM (see
REFERENCES


Fig. S1. The one-factor CFA model.
Fig. S2. The four-factor CFA model.
Fig. S3. The second-order CFA model.
Fig. S4. The bifactor CFA model.
Fig. S5. The four-factor ESEM model.

Fig. S6. The bifactor ESEM model.

Received 29 August 2023, Revised 15 November 2023, accepted 7 December 2023

© 2023 The Authors. Scandinavian Journal of Psychology published by Scandinavian Psychological Associations and John Wiley & Sons Ltd.