Early career burnout among nurses: Modelling a hypothesized process using an item response approach

J. Petter Gustavsson *, Lennart Hallsten, Ann Rudman
Division of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, SE-171 77 Stockholm, Sweden

What is known about the topic?

• Burnout is commonly regarded as one outcome of negative socialization among newly qualified human service professionals.
• Burnout is often accompanied by symptoms of low mood and intention to leave the profession, and may interfere with quality performance at work.
• Models of burnout often stress the multidimensional nature of the concept, although the various dimensions theoretically often are ordered sequentially, reflecting different phases.

What this paper adds

• Psychometric evidence indicates that early career burnout can be measured by a one-dimensional summative scale.
• Different levels of burnout on this scale reflect different phases in the burnout process.
1. Introduction

Early stages of professional socialization can be experienced as stressful, and despite several years of formal training, new professionals may initially feel inadequately prepared for their occupational role (Cherniss, 1980a,b; Duchscher, 2009). Moreover, values, ideals and expectations may clash with the harsh everyday reality at work (Maben et al., 2006; Mackintosh, 2006). Although induction or orientation programmes can provide novice professionals with some essential structure and guidance to develop competence in their new role (Price, 2009), they will inevitably be in a vulnerable position in the new work environment, sometimes being overwhelmed by frustration, self-doubt and feelings of failure (Duchscher, 2009). Cherniss has described this as an identity crisis among novice professionals, which in some cases leads to unsuccessful occupational socialization or early career burnout (Cherniss, 1980a,b). In cases where socialization develops into an erosion of engagement (i.e. early career burnout), energy turns into exhaustion, and involvement into cynicism (Cherniss, 1980a,b; Kramer, 1974; Mackintosh, 2006).

Within a longitudinal qualitative research programme, Cherniss (1980a,b) studied newly qualified human service professionals when they entered working life. Qualitative interviews were performed prospectively during the first years after graduation, and additional long-term follow-up interviews were performed after approximately 10 years (Cherniss, 1995). The initial aim of describing professional socialization within human service organizations proved to be a comprehensive outline of the process of early career burnout. Within this framework, the sources of the development of burnout were imbedded in the crisis of competence and a quest for respect that urged the novice to prove his or her worth and potential (Duchscher, 2008; Duchscher and Mironick, 2008; Kramer, 1974). Demanding, difficult or unmotivated clients, unsupportive colleges and organizational interference caused additional barriers to effective socialization and professional efficacy. Cherniss (1980a,b) found that some individuals coped with this stress by developing negative changes in attitude and behaviour. The early stages of negative attitude changes were characterized by changes in professional goals and aspirations, and a decline in responsibility and idealism in order to alleviate personal stress (see also Mackintosh, 2006). These attitudes were seen as a direct attempt to cope with the crisis of competence, difficult clients, unsupportive climate between colleagues, and organizational interference (Cherniss, 1980a,b). Temporarily these attitudinal changes offered some relief, and for some individuals this was the starting point for a transition towards a more sound professional identity. For others, additional attitudinal changes were needed in order to cope with the stress and the crisis of competence. In the later stages of negative attitudinal change (i.e. additional forms of coping, or perhaps consequences of inadequate coping), emotional detachment, cynicism and psychological withdrawal became the vehicle for dealing with stressors and strains. Used as a means of coping, these later changes promoted loss of commitment and in some cases the development of work alienation. Thus, at the heart of the Cherniss developmental model of early career burnout among human service professionals were the negative changes in attitudes towards clients, colleagues, accomplishments and work itself. Not only do these changes contribute to the development of burnout; rather, they constitute the erosion of involvement and commitment, i.e. burnout.

The concept of burnout has later been generalized beyond both early carrier issues and human professionals, and has also been defined as a syndrome comprising several core symptoms. When reviewing 25 years of research and theorizing on burnout (Schaufeli and Buunk, 2003; Schaufeli and Enzmann, 1998), Schaufeli found that over 100 symptoms of burnout have been listed in empirical research. However, these symptoms narrowed down to exhaustion and dysfunctional attitudes, or when using the scale labels from the most widely used instrument, the Maslasch Burnout Inventory (MBI): Exhaustion, Cynicism/Depersonalization and Personal Accomplishment. Although other dimensional models and instruments exist (Demerouti et al., 2001; Pines, 1993; Shirom, 2003), they generally converge in that they include the two core dimensions of exhaustion and dysfunctional coping (cynicism or disengagement). Schaufeli also concluded that in general the suggested developmental trajectories of burnout in general are in agreement when addressing precursors, core symptoms, and consequences. He proposed an integrative model of burnout (Schaufeli and Enzmann, 1998), postulating that burnout develops from an initial stress reaction due to a sharp contrast between an individual’s strong motivation and an unfavourable job environment. This initial stress reaction is often measured using an exhaustion scale. In addition, it is suggested that inadequate coping strategies (i.e. mental distancing) is crucial in the self-perpetuating process (as measured by a cynicism or disengagement scale), where both the initial threatening or taxing situation and the resulting stress reaction are handled ineffectively and exacerbated.

Taken together, different burnout instruments generally include the two main dimensions of fatigue (exhaustion) and dysfunctional coping (cynicism), and developmental models seem to converge, in that initial exhaustion develops, due to the application of dysfunctional coping (cynicism and disengagement), into burnout. In other words, exhaustion comes first, and if dysfunctional coping is applied, it progresses further. Interestingly, the theory outlined a developmental process, but the most common operationalizations reflect a two-dimensional syndrome, and in most burnout studies these two aspects are not combined into an assessment of a syndrome. Instead they are commonly analysed separately (Schaufeli and Buunk, 2003; Schaufeli and Enzmann, 1998). An interesting question is therefore whether it is possible to integrate the developmental model and a two-dimensional syndrome into a sequential-developmental measurement model. One possible method would be to apply the classic Guttman
method and scalogram analysis, as this method infers the order of both respondents and symptoms along a single latent continuum reflecting individual growth sequences (Nunnally, 1978). Thus, the method assumes that such a continuum is best operationalized with items indicative of different levels of (for example) the burnout process. Importantly, the responses to these items are thought to be completely determined by the location of the respondent and the item. For example, a person at a specific developmental stage in the burnout process will be characterized by both the symptoms typical of that stage, but also by the symptoms typical of the earlier stages. Thus, a person at an early stage of burnout may have had difficulties in regaining lost sleep, but does not need to show any signs of disengagement. In contrast, a person at a later stage of the developmental process will show increasing levels of disengagement but also express high levels of tiredness and problems regaining energy. Although Guttman scaling would be an appropriate method to evaluate a measure of the integrated model of the developmental process of burnout, the method is seldom recommended, since it does not take measurement error and randomness into account (Nunnally, 1978). Instead, several other probabilistic item response models may be used (Andrich, 1985; Cliff and Keats, 2003; Mokken, 1997) and among these the one-parameter IRT or Rasch model has gained widespread recognition for its measurement properties and usage in constructing and evaluating measures of developmental-sequential or hierarchical attributes. Thus, by using thispsychometric model, the overall purpose of this paper was to test a one-dimensional measurement model of early career burnout derived from the work of Cherniss.

1.1. Hypotheses

In this paper, the burnout process is divided – for illustrative purposes – into three different phases. The first phase (here called the pressure phase) concerns crisis and strain, reflecting a taxing work situation where demands drain resources, and more personal investments are made in order to cope, resulting in fatigue and distress. If these investments do not pay off and the situation further taxes one’s resources, this proceeds to the next phase (here called the frustration phase), characterized by higher levels of exhaustion, feeling drained and increase in defensive coping. The third and final phase is labelled the burnout phase, characterized by even higher levels of exhaustion and feeling drained, but adding withdrawal and emotional detachment as a means of escape. This sequential-developmental model is illustrated in Fig. 1. Three hypotheses will be addressed specifically:

**Hypothesis 1.** Burnout can be measured as a one-dimensional construct reflecting a sequential-developmental process where both items and individuals can be ordered sequentially.

**Hypothesis 2.** Items will specifically define unique phases along the sequential-developmental continuum, as illustrated in the hypothesized model in Figure 1.

**Hypothesis 3.** Different levels along the sequential-developmental measure will be associated with (a) external variables reflecting core symptoms of burnout (problems with energy and dysfunctional attitudes) and (b) variables previously found to be associated with different levels of burnout (such as low positive affect, work attitudes, and health).

2. Methods

2.1. Subjects

Data were taken from a prospective longitudinal study (the LANE study) that comprises three national cohorts of newly qualified nurses, followed from nursing education into their first years of working life (Rudman et al., 2009). For the purposes of the present study, the LANE cohort, consisting of 1155 nursing students who graduated in the autumn of 2002, was used. The present article includes data from a follow-up three years after their graduation from higher education. On this occasion, 933 responded to a survey, resulting in an 81% response rate. Median age was 31 years and 90% were women.

Permission to carry out the study was received from the Research Ethics Committee at Karolinska Institutet, Sweden, and all nursing students gave their written informed consent to participate in this study, well aware that they could terminate their participation at any time if they chose to.

2.2. Instruments

The LANE questionnaire comprises self-report instruments measuring background characteristics, psychological and physical health, social situation, healthcare utilization, sickness absence, social and psychological factors at work, as well as professional efficacy and quality aspects of the healthcare system (Rudman et al., 2009). The present paper utilizes items from the Oldenburg Burnout Inventory (Demerouti et al., 2001; Halbesleben and Demerouti, 2005). Importantly, OLBI per se is not being tested here, and if data had been available the hypothesis of a sequential-developmental measurement model of burnout could have been tested using other burnout instruments (for example, the Maslach Burnout Inventory). The Swedish version was translated from both the original German version and from an approved English version.
Back translation was performed and approved by the original constructor (Dahlin, 2007; Peterson, 2008). The 16 items in this inventory assess the two core dimensions of burnout, namely exhaustion and disengagement. Each item has a four-point response scale, ranging from 'Does not apply at all' to 'Applies completely'. The two scales each consist of eight items, where half of the items measure burnout phenomena and the other half the opposite pole of burnout (i.e., energy and engagement). Although the use of reversed worded items may be recommended to reduce acquiescence bias and response sets, psychometric analyses have found these reversed items problematic. For example, it has been shown that respondents may calibrate reversed worded items differently (Bode, 2004). In addition, it has been shown that when testing the dimensionality of measures of psychological symptoms that also comprise reversed worded items, there is a risk of finding that these items do not measure the same attribute (Betemps and Baker, 2004; Conrad et al., 2004). Perhaps more importantly, there is also reason to believe that the absence of psychological symptoms does not necessarily indicate psychological well-being (World Health Organization, 1947), in that operational definitions of psychological symptoms and psychological well-being do not capture different poles on the same dimension (Keyes, 2003). This discussion has been made explicit in the burnout literature (Maslach et al., 2001; Simpson, 2009) and empirical studies quite convincingly show that burnout and its proposed opposite pole, job engagement are qualitatively different, but negatively correlated phenomena (Duran et al., 2004; Schaufeli and Enzmann, 1998), were correlated with the burnout scale. The variables chosen included symptoms (e.g., depression, and stress) as well as negative work attitudes and poor job performance, and were taken from the Major Depression Inventory (Bech et al., 2001), the Nordic Questionnaire of Psychological and Social Factors at Work (Dallner et al., 2000), and from previous epidemiological studies on health among healthcare professionals (Dahlin, 2007; Peterson, 2008).

2.3. Data analysis

The first hypothesis stated above will be confirmed if psychometric analysis yields evidence of the presence of a one-dimensional internal structure where both items and individuals can be ordered according to their different levels on a (common logit) scale. In the present study, data were tested against a Rasch measurement (i.e. the partial credit) model (Andrich, 1988a,b; Wright and Mok, 2000). This one-dimensional measurement model is a mathematical model describing how data are supposed to behave according to the rules of probabilistic Guttman scaling (i.e. probabilistic order).

Table 1

| Item Code | Item Label | Item Fit | Item location | Thresholds
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Fit</td>
<td>χ²</td>
<td>p</td>
</tr>
<tr>
<td>1</td>
<td>Tired before work (e; pp)</td>
<td>-0.85</td>
<td>10.7</td>
<td>0.295</td>
</tr>
<tr>
<td>2</td>
<td>Devaluation of work (d; fp)</td>
<td>-2.69</td>
<td>15.5</td>
<td>0.078</td>
</tr>
<tr>
<td>3</td>
<td>Need longer time for rest (e; pp)</td>
<td>2.40</td>
<td>11.1</td>
<td>0.268</td>
</tr>
<tr>
<td>4</td>
<td>Mechanical execution of work (d; bp)</td>
<td>-0.12</td>
<td>6.0</td>
<td>0.744</td>
</tr>
<tr>
<td>5</td>
<td>Emotionally drained (e; fp)</td>
<td>-1.30</td>
<td>11.7</td>
<td>0.233</td>
</tr>
<tr>
<td>6</td>
<td>Lost inner relationship (d; bp)</td>
<td>-0.98</td>
<td>11.3</td>
<td>0.254</td>
</tr>
<tr>
<td>7</td>
<td>Fed up with work tasks (d; fp)</td>
<td>2.74</td>
<td>23.4</td>
<td>0.005</td>
</tr>
<tr>
<td>8</td>
<td>Worn out (e; pf)</td>
<td>0.64</td>
<td>9.6</td>
<td>0.388</td>
</tr>
</tbody>
</table>

Analysis of reduced item set (7 items)

| Item Code | Item Label | Item Fit | Item location | Thresholds
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tired before work (e; pp)</td>
<td>-0.39</td>
<td>9.8</td>
<td>0.370</td>
</tr>
<tr>
<td>2</td>
<td>Devaluation of work (d; fp)</td>
<td>-2.13</td>
<td>11.4</td>
<td>0.247</td>
</tr>
<tr>
<td>3</td>
<td>Need longer time for rest (e; pp)</td>
<td>1.77</td>
<td>15.2</td>
<td>0.085</td>
</tr>
<tr>
<td>4</td>
<td>Mechanical execution of work (d; bp)</td>
<td>0.54</td>
<td>9.1</td>
<td>0.423</td>
</tr>
<tr>
<td>5</td>
<td>Emotionally drained (e; fp)</td>
<td>-0.72</td>
<td>9.4</td>
<td>0.399</td>
</tr>
<tr>
<td>6</td>
<td>Lost inner relationship (d; bp)</td>
<td>0.17</td>
<td>12.3</td>
<td>0.195</td>
</tr>
<tr>
<td>8</td>
<td>Worn out (e; pf)</td>
<td>0.42</td>
<td>8.4</td>
<td>0.491</td>
</tr>
</tbody>
</table>

Data from the exploration sample, n = 393. Exhaustion (e) and disengagement (d) items are labelled and classified as belonging to one of the three hypothetical burnout phases (pp = pressure phase; fp = frustration phase; bp = burnout phase) outlined above (see also Fig. 1).

* Response category position given as thresholds; the first represents the threshold between the two lowest response categories, and the second and third represent the thresholds between the two middle response categories and the two highest response categories, respectively.
Second, the items locations along the logit scale were scrutinized, merely to confirm that items are spread out across the measured dimension. The intended increasing levels of severity across the response categories for each item also need to be reflected in the data. This will be reflected in that the locations of the response categories (i.e., estimated thresholds) for each item are spread across the scale and arranged in the same way as the manifest response categories are ordered. Another aspect of how person and item interacts is reflected in how well the items are targeted to the severity levels reported by the respondents. As the scale of the expected values is always centred on zero logits, calibrated against the item of average severity, the mean of the person’s locations could be compared with this value. If this mean value is close to zero (with an SD of about 1) the items are considered to be well targeted to the distribution of respondents in the sample.

Third, an additional test of uni-dimensionality was performed by submitting the matrix of residual correlations to a principal component analysis. The purpose of this test is to examine the structure not accounted for by the model. In RUMM this analysis starts by creating a new data matrix. In this matrix, each individual’s response to every item is substituted with the residual derived when subtracting the original score from the expected response (as assumed in the model). Correlations are then computed between the ‘new’ items (reflecting correlations among the items after the burnout dimension has been partialled out) and these correlations are further analysed in RUMM by a principal component analysis. If each item now loads highly in its own factor, and does not have a substantial loading on any other factor (i.e., on the first factor as these are ordered after how much variation they can explain), this will be interpreted as no other residual factor, indicating that multi-dimensionality is present in the data. Moreover, the test suggested by Smith (2002), where person locations based on subsets of items (as defined by results from the principal component analysis) are compared across all respondents, was also performed. A series of dependent t-tests were computed and the percentage of significant deviations was defined to not exceed 5%.

Finally, the psychometric analyses end with a fourth step: a test of whether the results are invariant across the confirmation sample. Specifically, the parameters estimated for the burnout scale in the exploration sample are used in order to compute new estimates in the confirmation sample. For each item, the two distributions of model-expected estimates are thereafter compared across successive groups representing increasing levels of burnout, using analysis of variance (ANOVA). Two types of deviations are evaluated and tested for. Uniform differential item functioning (uniform DIF) reflects that the estimates in the confirmation sample (for a given item) follow the
expected scores in the exploration sample but are consistently lower or higher across all severity levels. Non-uniform differential item functioning (non-uniform DIF) reflects that the estimates (for a given item) do not follow the expected scores but deviate with higher (and lower) scores on specific locations along the scale. This procedure for detecting measurement invariance was also used when testing for differential item functioning across gender and age groups. A median split was used to dichotomize age into two groups (up to 31 years vs 32 years and above).

All these psychometric tests concern aspects of the internal structure of the instrument, and further tests are needed to confirm that the instrument also relates to other (external) variables in an expected direction. A test of whether the hierarchical structure modelled by the Rasch analysis also relates to the hypothetical structure outlined above (and illustrated in Figure 1) was performed in order to test Hypothesis 2. Item and threshold locations were graphically plotted against the hypothesized phases. Furthermore, to test the third hypothesis, external variables commonly found to be concomitants and consequences of burnout were correlated with the proposed burnout measure. In this study the external variables were classified as related work aspects (job stressors, work attitudes and job performance), mood (low positive affect such as: not feeling inspired, not feeling optimistic, have lost interest, lacking energy, feeling slow and sluggish, difficulty in concentrating, difficulty in making decisions), and health (poor SRH, physical symptoms of tension and pain—i.e. muscular tension/aches, headache, neck pain, back pain). Correlation coefficients were computed and the prevalence of increasing levels of associated symptoms and attitudes across these phases was tested by the \( \chi^2 \) statistic. Logit scores for items and respondents, i.e., the expected scores estimated under the assumption of model fit were interpreted, and cut-off scores were decided upon approximating the proposed phases.

3. Results

The results are presented under two headings, corresponding to the psychometric evaluation scrutinizing evidence for the scales internal structure (i.e., test of the first hypothesis), and the evaluation of evidence concerning association with external variables (i.e., test of the second and third hypotheses).

3.1. Psychometric analysis: test of the first hypothesis

Psychometric analysis begins by comparing observed values against expected values, assuming that model fits the data. In this analysis, an overall test is performed of whether the total item pool of eight items (as responded to by the 319 subjects in the exploration sample) complies with the expectations of the uni-dimensional Rasch model. The difference between these model-expected values and observed values, i.e. the residuals, are scrutinized and tested using the global statistical test comparing these differences along the total scale. This test reveals that data do not fit the model \( (\chi^2 = 99.20; \text{df} = 72; \ p = 0.018) \). Furthermore, if data fit the model, both item and respondent residuals should have a distribution with a zero mean and 1 standard deviation. A comparison with the expected mean value of 0 (with 1 SD) of the residuals reveals that there is considerably more variation \( (M = 0.02; \ SD = 1.86) \) among the item residuals than expected. Thus, the problem of misfit may be due to the presence of one or several misfitting items. Fit values of the eight items are presented in Table 1. Only item number 7 had a fit value over 2.5 and a statistically significant deviation between model-expected values and observed data. In addition, good fit will also be indicated by a response pattern reflecting consistent use of the response options, and the results show that the three estimated thresholds for each item were in all cases arranged in the same order as the manifest response categories (Table 1).

One item (no. 7, Fed up with work tasks) captures negative feelings about work tasks. Although negative feelings about certain work tasks are thought to be part of the middle phases of burnout, this item does not actually fit. The item was deleted and the analyses started from the beginning with the seven remaining items. Descriptive data and overall tests of fit of the new set of 7 items are given in Table 1. The fit has improved considerably, reflected in both a non-significant \( \chi^2 \) value \( (\chi^2 = 75.74; \text{df} = 63; \ p = 0.130) \) and lower standard deviations of item fit residuals \( (M = 0.01; \ SD = 1.21) \). The reliability estimate was above 0.84, indicating acceptable internal consistency. Locations of items and thresholds on the suggested burnout continuum are given in Table 1, along with item fit statistics. All items now fit the model and the item thresholds reflect consistent use of the response categories across all items.

A test of whether each respondent fits the hypothesized pattern was also evaluated. Again, the residuals were used and a fit value for each respondent was calculated. The results from this analysis revealed that only three respondents substantially deviated from the model with fit values higher than 2.5. These three respondents all scored lower than expected on items reflecting tiredness after work and need of recovery time (all fit values between 2.5 and 2.9). Another aspect of how person and item interact is reflected in how well the items are targeted to the severity levels reported by the respondents. The mean value for the respondents' locations was \( (1.1; \ SD = 1.5) \), which indicated that the items were not well targeted to respondents with lower scores, i.e. respondents who had no problems with early career burnout. In total, 16 respondents had the lowest possible sum score, and the items cannot measure differences among these. As the numbers of respondents that deviated substantially from the expected response pattern were so few, the one-dimensional model of developmental phases was confirmed also for the respondents.

All these tests reflect that the data adhere well to the Rasch measurement model, indicating the measurement of a one-dimensional construct. However, a further test of the correlational structure for the residuals may be warranted, and in such a test the varimax-rotated components should exhibit a pattern where each and every item loads
Data from the exploration sample, n = 393.

non-significant ($\chi^2 = 72.26$; df = 63; $p = 0.196$). Second, the test of whether the item parameters are equivalent across both subsamples (presented in Table 3) showed no significant differences, indicating that neither uniform DIF nor non-uniform DIF was present. Thus, the test confirmed that the seven items function invariantly across the two samples.

Also in Table 3, tests of invariance across gender and age groups are presented. The results indicate a presence of uniform DIF for several items, suggesting that different gender or age groups show consistent systematic difference in endorsement to specific items (for example consistently high or low endorsements) along the whole burnout continuum. For example, the largest gender difference was found for item no. 4 (mechanical execution of work) and an inspection of data suggests that males are more likely to endorse this item across all levels of burnout. The largest age difference was found for item no. 3 (need longer time for rest) and an inspection of data suggests that younger subjects are less likely to endorse this item across all levels of burnout. The practical implications for these differences were further evaluated by testing for mean differences (between gender and age groups) and computing correlations, using the original scale and an adjusted scale where the non-equivalent items have been calibrated separately in each subgroup (for gender item no. 4, 6 and 2 were identified as non-equivalent in a stepwise procedure and calibrated separately; for age item no. 3, 1 and 2 were

Table 2
Principal component analysis of the residuals showing the total varimax rotated matrix.

<table>
<thead>
<tr>
<th>Item/Component:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Mechanical execution</td>
<td>0.951</td>
<td>-0.121</td>
<td>-0.115</td>
<td>-0.061</td>
<td>-0.158</td>
<td>0.020</td>
<td>-0.194</td>
</tr>
<tr>
<td>1. Tired before work</td>
<td>-0.125</td>
<td>0.938</td>
<td>-0.144</td>
<td>-0.160</td>
<td>-0.129</td>
<td>-0.156</td>
<td>-0.134</td>
</tr>
<tr>
<td>2. Devaluation of work</td>
<td>-0.108</td>
<td>-0.131</td>
<td>0.967</td>
<td>-0.089</td>
<td>-0.087</td>
<td>-0.042</td>
<td>-0.136</td>
</tr>
<tr>
<td>5. Emotionally drained</td>
<td>-0.059</td>
<td>-0.149</td>
<td>-0.092</td>
<td>0.960</td>
<td>-0.136</td>
<td>-0.023</td>
<td>-0.159</td>
</tr>
<tr>
<td>8. Worn out</td>
<td>-0.169</td>
<td>-0.134</td>
<td>-0.099</td>
<td>-0.151</td>
<td>0.927</td>
<td>-0.240</td>
<td>-0.058</td>
</tr>
<tr>
<td>6. Lost inner relationship</td>
<td>0.019</td>
<td>-0.159</td>
<td>-0.049</td>
<td>-0.027</td>
<td>-0.233</td>
<td>0.938</td>
<td>-0.190</td>
</tr>
<tr>
<td>3. Longer times for rest</td>
<td>-0.235</td>
<td>-0.155</td>
<td>-0.173</td>
<td>-0.196</td>
<td>0.196</td>
<td>-0.065</td>
<td>-0.217</td>
</tr>
</tbody>
</table>

Table 3
Test of differential item functioning (DIF). Test of main (uniform DIF) and interaction (non-uniform DIF) effects.

<table>
<thead>
<tr>
<th>Item, Component</th>
<th>DIF across samples</th>
<th>DIF across gender</th>
<th>DIF across age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Uniform-DIF</td>
<td>Non-Uniform DIF</td>
<td>Uniform-DIF</td>
</tr>
<tr>
<td>MS F Prob MS F Prob MS F Prob MS F Prob MS F Prob MS F Prob</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tired before work</td>
<td>0.48 .62 .432</td>
<td>1.19 1.54 .130</td>
<td>1.29 1.59 .208</td>
</tr>
<tr>
<td>Devaluation of work</td>
<td>0.01 .02 .908</td>
<td>.35 .49 .879</td>
<td>0.25 .37 .545</td>
</tr>
<tr>
<td>Need longer time for rest</td>
<td>1.15 .17 .682</td>
<td>1.25 1.33 .216</td>
<td>14.94 15.92 .001</td>
</tr>
<tr>
<td>Mechanical execution of work</td>
<td>0.53 .58 .445</td>
<td>1.13 1.23 .274</td>
<td>15.14 17.49 .001</td>
</tr>
<tr>
<td>Emotionally drained</td>
<td>0.06 .08 .781</td>
<td>.53 .71 .697</td>
<td>3.78 4.49 .029</td>
</tr>
<tr>
<td>Lost inner relationship</td>
<td>0.01 .01 .992</td>
<td>1.20 1.20 .289</td>
<td>5.42 6.32 .012</td>
</tr>
<tr>
<td>Worn out</td>
<td>0.01 .01 .962</td>
<td>1.60 1.78 .068</td>
<td>1.46 1.66 .198</td>
</tr>
</tbody>
</table>

Data from the total sample.
identified as non-equivalent in a stepwise procedure and calibrated separately). First, mean differences between the two age groups were found to be approximately the same, regardless of whether the scale was adjusted for the non-equivalent items or not. Younger subjects were found to have significantly higher levels of burnout (effect sizes, Cohen’s d, of 0.34 and 0.34 when tested with the original and adjusted scale, respectively). In contrast, when using the adjusted scale for the gender comparison, male subjects were found to have significantly lower burnout scores than the female subjects (effect sizes of 0.14 and 0.41 when tested with the unadjusted and adjusted scale, respectively). Thus, these analyses suggest that at least for the small subgroup of men, items reflecting higher levels of burnout may need to be calibrated separately. Second, extremely high correlations between the original burnout scale and the adjusted scales were found \( r = 0.993 \) between the original and the gender-adjusted scales, and \( r = 0.998 \) between the original and the age-adjusted scale), indicating that these calibrations do not have any impact when measuring individual differences in burnout levels (all further computation below is therefore on the unadjusted scale).

3.2. Comparing the psychometric model with the hypothetical model: test of the second and third hypotheses

Up to this point, all testing indicates (with the possible exception of the small male subgroup) that the data adhere to a sequential-developmental, i.e., a probabilistic Guttman-like, dimensional structure. But, is this established pattern, although evidently psychometrically sound, really indicative of the developmental process of early career pattern, although evidently psychometrically sound, really indicative of the developmental process of early career pattern, although evidently psychometrically sound, really indicative of the developmental process of early career pattern, although evidently psychometrically sound, really indicative of the developmental process. In Fig. 2, the items are rank ordered from top to bottom, indicating increasing severity (i.e., locations along the continuum). In addition, item threshold locations are given, reflecting the partitioning of the continuum, from right to left, indicating increasing severity. First, it can be seen that items measuring various levels of pressure come first (tired, worn out, need more rest), followed by items dealing with frustration (talk about work in a derogatory manner feel emotionally drained), followed by items capturing the last phase comprising elements of withdrawal and detachment (have lost interest, approach work mechanically). Generally, moving from left to right, the locations appear to increase monotonically. In particular, the locations reflect different, but overlapping, segments of the displayed continuum. Thus, the estimated structure seems to support the hypothesis of a developmental model of burnout consisting of (at least) three broad phases. Secondly, when trying to interpret the relative locations of the thresholds (presented in Fig. 2), and relate these to the phases proposed, it can be seen that fatigue problems start at around 0.6 on the scale and reach a maximum at around 1. Cynicism and feeling emotionally drained start before 1 and reach a maximum level after 2. Finally, withdrawal and detachment start after 1 and reach a maximum level after 2.5. If we use these points along the continuum as approximate cut-offs, we find that 62% are classified as free from symptoms, whereas 24%, 9% and 5% are classified as pressured, frustrated, and burned out, respectively.

As a test of the third hypothesis, the association between the sequential structure and external variables will be evaluated (results presented in Tables 4 and 5). Table 4 shows correlations between the burnout measure and work-related variables such as prevalence of job stressors and spillover, feelings towards work and intention to quit, as well as satisfaction with work performance and accomplishment. Prevalence of such problems across the four different sequential phases is also presented. The largest correlations indicate that present stressors in the work situation – feelings of indifference towards work and intention to quit – are the external variables most related to the sequential-developmental burnout variable. For all variables, the prevalence increases monotonically across the four different phases.

In Table 5, the variables related to low positive affect are correlated with the burnout measure. The largest correlations are found for lack of energy and loss of interest in everyday activities. Again, the prevalence of various mood states (i.e. low positive affect) increases monotonically across the burnout continuum. Also in Table 5, comparisons are made against health-related variables. The largest effect size for these health variables was found for self-rated health, although this effect was not as high as those reported for work attitudes and low positive affect. The prevalence of health problems increases monotonically across the hypothesized sequential dimension.

![Fig. 2. Item hierarchy of burnout items. Item location and response thresholds given on the same scale.](image)
4. Discussion

In a national sample of registered nurses, three years after graduation, items reflecting varying levels of exhaustion reactions and negative attitudes towards work were found to adhere to a one-dimensional measurement model. This model of early career burnout among nurses was interpreted as indicating a sequential-developmental model, where there is a progression from initially increasing levels of exhaustion into burnout, due to dysfunctional coping (cynicism and disengagement). Or to put it another way, the measurement model indicated that different levels of burnout reflect that exhaustion comes first in the burnout process, and then develops further if dysfunctional coping is applied. Moreover, other measures of problems with energy and engagement monotonically paralleled the measured continuum. These results support the original model of early career burnout that originated from the longitudinal grounded-theory investigations by Cherniss (1980a,b). Although there is other research supporting the Cherniss framework (Burke and Greenglass, 1995, 1989; Burke et al., 1984), studies have mostly confirmed the importance of crucial predictors of burnout as outlined in his model. (For a review, see Burke and Richardsen, 2001). In contrast to such studies, the present study focused exclusively on the burnout phenomena outlined by Cherniss (1980a,b), and by applying a psychometric approach modelling sequential levels of burnout, confirmed the structure of the phenomena as suggested by his model.

Does other research using other methods support this sequential-developmental model? On the basis of a thorough review of theoretical models and empirical studies of the burnout process, Taris et al. (2005) (using data from several longitudinal studies) modelled different causal relationships between components in the burnout process. Using the longitudinal method of cross-lagged correlations they found evidence for several causal mechanisms. In their empirical study, they presented support for the notion that the burnout process starts with increasing levels of exhaustion, and by using disengagement as a coping strategy, levels of both exhaustion and disengagement increase even further. Specifically, Taris and co-workers found significant lagged correlations, reflecting that higher levels of exhaustion longitudinally predicted higher levels of disengagement. At the same time, higher levels of disengagement predicted further increases in exhaustion.

Furthermore, the results presented may be in line with research specifically addressing the process of becoming a
professional nurse. Similar attitudinal changes of disengagement that are observed in the burnout process are also to some degree recognized among those undergoing socialization into a range of professions, such as nurses, teachers, psychotherapists and lawyers (Cherniss, 1980a,b; Watson et al., 1999). The results presented here may therefore in some instances partly reflect a negative socialization process, with overall shortfall of humanistian standards and ideals, as an effect of the efforts of new nurses to acquire a professional role (Mackintosh, 2006; Morse et al., 1992; Watson et al., 1999). In a longitudinal qualitative study, Mackintosh (2006) found that many new graduates established a degree of “emotional hardness” and cynicism during the first 6-9 months of practice, which affected the way they cared about patients. New nurses regarded the ability to “switch off” as a crucial element to be able to cope with work, to preserve themselves from the distressing nature of some practice experiences, and to demarcate home and working life (Mackintosh, 2006). However, one interesting question that remains largely unanswered, is why some individuals allow attitudinal changes such as cynicism in themselves, whereas others maintain a caring ethos despite exposure to poor role models and daily pressures in the work environment (Mackintosh, 2006; Morse et al., 1992). Morse et al. (1992) suggest exhaustion or burnout as one possible explanation when nurses emotionally separate themselves, become less sensitive to the suffering cues of patients and in the worst cases completely detached. Morse also suggests that the mechanism of detachment may have several functions. Detachment can for instance serve as a necessary means for the caregiver to overcome the stress caused by the patients suffering, when his or her own feelings would be too painful to handle. In this case, detachment allows the caregiver to carry on working since providing disengaged care implies a reduced effort. On the other hand, in some cases, when for example unpleasant or serious situations occur it can be beneficial for the patient that the caregiver can conceal his/her own feelings of fear or disgust. Despite the sometimes beneficial outcomes of detachment, patients are aware when they are not being treated as persons or when their needs are ignored (Coyle, 1999; Morse et al., 1992; Strauss et al., 1982). The implications for patients when they are treated as objects rather than individuals are described at length in the research of patient dissatisfaction (Coyle, 1999; Strauss et al., 1982; Williams and Calnan, 1991). One of the most significant factors in relation to patient satisfaction with care is interpersonal manner, i.e., how care providers interact personally with patients and the degree to which they show humanity, courtesy, respect, reassurance and friendliness (Ware et al., 1983). The disengagement may have implications for caregivers as well. They may lose emotional insight and the coping mechanism that started out as a way of being able to keep working, escalates and ends up in exhaustion and ineffective management of stress, resulting in still more pressure and strain (Cherniss, 1980a,b), as was also found in Taris et al.’s study. However, when interpreting the present results in the light of the transition and socialization process for nurses, one should keep in mind that the process was primarily described throughout the first year after graduation (Duchscher, 2008; Mackintosh, 2006), in comparison with the burnout assessment of this study, which was made three years after graduation.

Although high levels of exhaustion without accompanying levels of disengagement can be found in the sequential-developmental burnout process (but not the other way around, i.e., high levels of disengagement without high levels of exhaustion), this perspective and the results presented here are in contrast with studies where separate cut-off points on the exhaustion and disengagement scales have been used to demarcate the burnout syndrome. In one such study, the total population of all employees in a county council was assessed and a group comprising about 12% of the total sample was characterized only by high levels of disengagement (Peterson et al., 2008a,b). However, the size of this group varied considerably across different occupations (from about 6% to 27%), and was found to be lowest among nurses, paramedics and physicians. In the present study, the item responses from a small part of the sample did not adhere to the sequential-developmental measurement model (i.e., their responses were found to deviate substantially from the probabilistic Guttman structure). These three respondents all scored lower than expected on items reflecting tiredness after work and greater need for recovery time, indicating a response pattern where high scores have only been obtained on disengagement items. This small number of individuals can obviously not be used as evidence for the existence of high levels of disengagement without high levels of exhaustion; however, if the prevalence of such respondents increases with more experience in working life, it will be interesting to track the unique trajectories for this particular group. For example, it would be of great interest to know whether these individuals have a recent history of burnout, and whether the disengaged coping style may be a recessive or an indicator of burnout progressing into remission. Thus, longitudinal studies of these newcomers may further validate the sequential-developmental model of early career burnout and perhaps equally importantly shed light on the probable phases of recovery.

The psychometric modelling presented in this paper used an item pool, comprising eight burnout items from a previously used burnout inventory (Demerouti et al., 2002; Halbesleben and Demerouti, 2005). In the process, one item (no. 7, Fed up with work tasks) was found to deviate significantly and substantially from the proposed model, and did not fit together with the other items on the suggested burnout continuum. The content of the item captures negative feelings about work tasks, and the misfitting item was hypothesized to be an indicator of the frustration phase of burnout. The reason for misfit may be interpreted as a reflection of how nurses think about their profession in contrast to their current position. Or as cited in a recent study on implicit and explicit factors that contribute to newly qualified nurses’ dissatisfaction and distress within hospital nursing: “I love being a nurse, but I hate my job” (Duchscher and Myrick, 2008). Thus, an item which focused more on emotional reactions related to the general work situation (instead of the profession), might
have performed better in our measurement model. However, the original item may work when assessing early career burnout in other occupational groups. Future work may well try to include additional items, and perhaps most importantly, items assessing beginning problems of exhaustion may be most warranted, as it was found that a proportion of the individuals exhibited extremely low scores and no differentiation could be made between them. Importantly, the originally two-dimensional measurement model of OLBI was not subjected to test in this paper. Neither was the framework often used in OLBI studies, i.e., the demand-resources model of burnout (Demerouti et al., 2001) used when formulating hypotheses about the developmental-sequential structure of burnout. Instead, only the pool of burnout items from OLBI has been used and subjected to a test of the one-dimensional model of burnout operationalized from the literature on early career burnout (Cherniss, 1980a,b). Thus, future work may also test the hypothesis of a sequential-developmental measurement model of burnout using other burnout instruments (for example, the Maslach Burnout Inventory).

The test of data from registered nurse three years after graduation responding to items from a burnout inventory against the requirement of a Rasch measurement model revealed that data adhered rather well to the model. Model fit is however always relative (Smith, 2000) and although the present analyses show indications of good model fit, there is certainly room for improvement. One such improvement influencing model fit deals with the issue of how well the items are targeted to reflect the symptom levels of the respondents. As described above, inclusion of items measuring early symptoms of exhaustion would certainly increase model fit. Furthermore, the principal component analysis of the residual data (original data where model-expected data had been partialled out), did not result in the finding of a substantial ‘second’ factor not accounted for by the extracted one-dimensional model. Such a finding would have strongly questioned our interpretation of the presence of a one-dimensional model. Again, the good fit is relative and there are residual correlations not accounted for by the model. These residual correlations approach zero when scrutinizing associations among items within each of the three phases. However, statistically significant negative residual correlations were found (as reflected in the significant loading found on the first component) between certain items reflecting the pressure phase and some items reflecting the burnout phase. Thus, to a minor extent, residual exhaustion variance is negatively associated with residual disengagement variance. The impact of this latter finding was evaluated by the procedure suggested by Smith (2002), comparing person locations derived from the different subsets of items against the total scale, and again the presence of a one-dimensional model was confirmed.

The present study draws heavily upon the grounded-theory studies by Cherniss (Cherniss, 1989, 1980a,b). Studies that were originally set up to study early career socialization, discovered the burnout process, and the framework describing the process, and the important predictors of progress are still among the most vital theories in the field. An example of this can be seen in how the important components in the Cherniss framework are all included as crucial parts in the integrative definition and model suggested by Schaufeli (Schaufeli and Buunk, 2003; Schaufeli and Enzmann, 1998). Our psychometric study adds to these studies, confirming aspects of the model outlined by Cherniss. However, it must be remembered that the items analysed in this study were not originally constructed to match the Cherniss model. In addition, this study focused only on modelling an outcome, and as such applied a rather limited aspect of the total model. Future studies of early career burnout may well informed by the theoretical model that inspired the present research. Especially prospective research may further inform us about the validity of the model, and should extend the original model by establishing the magnitude and relative importance of different predictors of burnout as well as its consequences.

Conflict of interest
None.

Funding
Grants were received from AFA Insurance. No other participation (than funding).

Ethical approval
Permission to carry out the study was received from the Research Ethics Committee at Karolinska Institutet, Sweden, and all nursing students gave their written informed consent to participate in this study, well aware that they could terminate their participation at any time if they chose to.

References


