DEVELOPMENT OF A MODIFIED EXHAUSTION STROOP TASK BASED ON THE SMBQ

Elin Stenudd, Rebecka Tholerud
Abstract

To explore the application of a cognitive test useful in screening of Exhaustion Disorder (ED), this study aimed to develop a Modified Stroop Task for ED. Participants’ scores on measures of burnout, anxiety, depression and stress were compared with performance on the task. The sample consisted of 56 participants (M=25.3 years), 30 women and 22 males, with the majority enrolled in fulltime higher education. The task consisted of five blocks with 100 stimuli in each block. The task was administered on a computer; upon completion of the task participants completed questionnaires measuring levels of burnout, anxiety and depression, as well as perceived stress. Raw scores on the task were calculated using a format comprised of both response times and number of errors. Data were analysed by comparing mean scores on the Stroop blocks using one-way repeated measures ANOVA. Additionally, median splits were undertaken on the scores on the questionnaires to enable between group comparisons using MANOVAs. The results showed no significant differences between high and low scores on the questionnaires and performance on the task. A Stroop interference effect was observed and there was a tendency that high exhausted participants performed worse than less exhausted participants. In conclusion the Modified Stroop Task offers some promise, although as of yet it cannot be assumed to be an accurate modification of an Emotional Stroop Task, due to limitations regarding the method; preliminary findings may serve as a pilot test for future research.

Keywords: stroop, exhaustion, burnout, emotional stroop, attentional bias

Abstrakt


Nyckelord: stroop, utmattning, utbrändhet, emotionell stroop, uppmärksamhetsbias
Development of a Modified Exhaustion Stroop Task based on the SMBQ

According to statistics from the Swedish Social Insurance Agency (2015), rates of long-term sick leave are increasing. Between 2010 and 2015, psychiatric diagnoses increased by 119% and are now the most common cause of sick leave for both men and women (Försäkringskassan, 2015). Similar patterns of increased mental illness have also been observed in other Western societies, where 20% of the working-age population is expected to suffer from a mental disorder (OECD, 2012). Stress-related mental illnesses account for 49% of this increase in Sweden, and are therefore the most common psychiatric diagnoses contributing to sick leave. Exhaustion Disorder (ED) is one of the diagnoses captured under this category, and is more likely than other diagnoses to lead to reduced working ability (Försäkringskassan, 2015). In light of the increasing incidence of mental illness and ED in particular, the development of alternative screening methods for use in clinical settings to capture this diagnosis appears warranted. Therefore, the aim of the present study is to develop a cognitive test for the purpose of examining potential differences in executive functioning between individuals scoring either high or low in exhaustion.

In the research literature, exhaustion and burnout are often used interchangeably as synonyms, with burnout appearing to be the most common concept. Although, the conceptualization of burnout has varied in research over the past two decades, in 2001 Maslach and colleagues defined burnout as a prolonged response due to emotional and interpersonal stressors on the job and advanced a multidimensional theory of burnout that has guided a large body of research. The theory includes three dimensions: emotional exhaustion, depersonalization, and reduced personal accomplishment. Emotional exhaustion refers to a decreased amount of positive emotions directed at the people one cares about, while depersonalization is characterized by an increase in cynical and callous attitude towards them. Reduced personal accomplishment refers to increased feelings of being ineffective in work settings (Maslach, Schaufeli, & Leiter, 2001).

An alternative definition gaining a great deal of interest in the research literature was put forward by Melamed, Kushnir, and Shirom (1992). They defined burnout as a syndrome consisting of three main components: physical fatigue, emotional exhaustion, and cognitive weariness (Melamed et al., 1992). Although these two definitions of burnout differ, they both emphasize exhaustion as a key component. Studies have shown that people fulfilling the criteria for ED, also score high on burnout scales (Glise, Ahlborg, & Jonsdottir, 2012; 2014). Additionally, a recent literature review on burnout (Grossi, Perski, Osika, & Savic, 2015) argues that ED is the most valid operationalization of clinical burnout. The Swedish Board of Health and Welfare (2003) has also proposed criteria to be used in clinical settings in order to diagnose ED (see Table 1).
Diagnostic criteria for exhaustion disorder in Sweden

A) Physical and mental symptoms of exhaustion with minimum two weeks duration. The symptoms have developed in response to one or more identifiable stressors, which have been present for at least 6 months.

B) Markedly reduced mental energy, which is manifested by reduced initiative, lack of endurance, or increase of time needed for recovery after mental efforts.

C) At least four of the following symptoms have been present most of the day, nearly every day, during the same 2-week period:
   1. Persistent complaints of impaired memory.
   2. Markedly reduced capacity to tolerate demands or to work under time pressure.
   3. Emotional instability or irritability.
   4. Insomnia or hypersomnia.
   5. Persistent complaints of physical weakness or fatigue.
   6. Physical symptoms such as muscular pain, chest pain, palpitations, gastrointestinal problems, vertigo, or increased sensitivity to sounds.

D) The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

E) The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hypothyroidism, diabetes, infectious disease).

F) The stress-related disorder does not meet the criteria for major depressive disorder, dysthymic disorder, or generalized anxiety disorder.

Stress-related mental illness, ED in particular, appears to be interrelated with other manifestations of mental illness. Among a Swedish sample of primary health care patients, individuals with high levels of stress frequently present symptoms of ED, depression, and anxiety (Wieger, Hange, Björkelund, & Ahlborg, 2015). Several studies support the notion that high levels of stress, burnout, and depressive disorders are clearly related and have overlapping conditions (Ahola et al., 2005; Hakanen & Schaufeli, 2012; Wang, 2005). A recent review noted that the conceptual distinction between burnout and depression is insufficient, and that knowledge regarding how burnout and clinical depression differ is unclear and should be further studied (Bianchi, Schonfeld, & Laurent, 2015). Additionally, some studies have found stress and burnout to be associated with anxiety (Ding, Qu, Yu, & Wang, 2014; Zhou et al., 2016). Thus, the presence of other manifestations of mental illnesses appears to be associated with exhaustion. However, other researchers argue that they are separate states that share several symptomatic characteristics (Iacovides, Fountoulakis, Kaprinis, & Kaprinis, 2002). Grossi (2015) posits that despite a growing body of literature on physiological correlates of burnout, there is no consensus on this matter.

Research examining exhaustion has been univocal in the sense that exhaustion is negatively related with job performance (Bakker, Van Emmerik, & Van Riet, 2008; Taris, 2006). Many theoretical explanations for this relationship have emerged (e.g., Job demand-Resources theory by Bakker & Demerouti, 2017), but there is limited consensus on the
underlying mechanisms. In particular, recent research has consistently highlighted that exhaustion is associated with cognitive deficits in executive functions, attention as well as episodic memory, and working memory (Deligkaris, Panagopoulou, Montgomery, & Masoura, 2014; Grossi et al., 2015).

Patients who seek medical care for stress-related mental illness frequently report cognitive impairments as the most pronounced symptoms (Jonsdottir et al., 2013). Deligkaris and colleagues (2014) propose that the components regarding attentional control functions are vulnerable to chronic stress, since they are more demanding and complex. In particular, one study found that people with higher levels of exhaustion performed less well on cognitive tasks requiring a high level of executive control, with longer response times as well as more errors, than those with lower levels of exhaustion (Diestel, Cosmar, & Schmidt, 2013).

Earlier research has identified many alternative factors that explain performance variations on cognitive tasks among individuals; one explanation that has been consistently highlighted in research is the effect of age. It is generally accepted that performance on cognitive tasks, such as the Stroop task, diminishes with age (e.g., Milham et al., 2002; Verhaeghen & Cerella, 2002). Various theories are proposed to explain this observed effect with the position that there is a deficit in processing speed; this is supported by research demonstrating a general slowing effect with age (Verhaeghen & Cerella, 2002). Others argue that deficits are due to age-specific changes in the prefrontal cortex (Milham et al., 2002). This contention is supported by research of neural activity, where the brain activity of younger people was compared with older participants’ during a Stroop task. Specifically, one study observed age-related decreases in the responsiveness of prefrontal structures supporting attentional control (Milham et al., 2002).

Cognitive impairment is common among patients with stress-related mental illness, and exhaustion in particular (Diestel et al., 2013; Deligkaris et al., 2014; Ellbin, Engen, Jonsdottir, & Nordlund, 2017), therefore the development of a neuropsychological test that measures cognitive impairment associated with ED appears to be warranted. However, limited research has examined which tests are most effective. A recent study attempting to address this issue measured performance on several cognitive tasks among patients with stress related exhaustion; performance on a Stroop test was observed to be significantly lower in patients with ED compared to controls (Ellbin et al., 2017).

The Stroop task was originally designed by Stroop (1935) as a test aiming to measure cognitive interference during the presentation of incongruent ink-colors displaying specific color-words (e.g., the word blue presented in a red color). The Stroop test has been extensively used in research and clinical settings for decades and is generally acknowledged as a cognitive test measuring cognitive abilities such as executive functions and attention (MacLeod, 1991; Rabin, Barr, & Burton, 2005). In particular, it has widely demonstrated its ability to differentiate between populations comprised of psychological conditions that affect cognition (Bowes, Stroman, & Garcia, 2012; Dikmen, Machamer, Winn & Temkin, 1995; Hallion, Folin, Assaf, Goethe, & Diefenbach, 2017; Homack & Reccio, 2004; Lansbergen, Kenemans, & van Engeland, 2007).

The Stroop task was first administered on sheets of paper where color-words were written in non-corresponding ink-colors; the respondent completing the task was instructed to verbally name the ink-color of each word in sequential order (Stroop, 1935). The hypothesis was that response times would increase when the color-words were presented with non-corresponding ink-color, and this would measure the extent of the interference of color stimuli upon reading words (Stroop, 1935). Later, MacLeod (1991) undertook a review of a vast number of studies published on the Stroop Effect, in summarizing the findings and evaluating the major theories used to explain the experiment, he proposed “the automaticity of reading
theory”. This theory argues that reading is more automatic than naming a color, which then requires more attention and thus increases response time for stimuli where reading and color-naming interfere (MacLeod, 1991).

As a result of the vast use of the original Stroop task, there has been a wealth of studies designing and evaluating various modified Stroop tasks (Cisler & Koster, 2010; Phaf & Kan, 2007). In these modified versions, different types of words are displayed in varying colors with the same instructions as the classic Stroop task (Cisler & Koster, 2010). For example, increased response times and more errors have been observed in experiments using threat-related words compared to neutral words displayed in different blocks; this finding has been explained as attentional bias (Frings, Englert, Wentura, & Bermeitinger, 2010). This paradigm has come to be known as the Emotional Stroop Task; the paradigm includes adapted versions of the task for different anxiety disorders, which appears to be sensitive to all disorders. In particular, attentional bias appears within Emotional Stroop Tasks developed for GAD, PTSD, specific phobia, panic disorder and OCD (Becker, Rinck, Margraf, & Roth, 2001; Buckley, Blanchard & Hickling, 2002; Cisler et al., 2011; Constantine, McNally & Hornig, 2001; Foa, Ila, McCarthy, Shoyer, & Murdock, 1993). Cisler and Koster (2010) therefore argue that attentional bias is likely present in all anxiety disorders. Many researchers have developed theories as to why attentional bias towards threatening words occurs (e.g., Beck & Clark, 1997; Phelps & LeDoux, 2005), but there remains a lack of theoretical agreement upon several questions regarding attentional bias and its components (Cisler & Koster, 2010). However, there is an accordance in all theories on the assumption that threat detection is an automatic process underlying a facilitated attention to threat (ibid).

In review of previous research, the present study aims to develop a Modified Stroop Task for exhaustion for the purpose of examining potential differences in performance between individuals scoring either high or low in exhaustion. The Modified Stroop Task will consist of different blocks in order to explore the influence of exhaustion-related words compared to neutral words. It will also include blocks based on the original Stroop task in order to compare the Stroop effect between these groups, based on levels of exhaustion. In consideration of previous research, it is expected that an attentional bias towards exhaustion-related words will be observed; specifically, highly exhausted individuals will be predisposed towards longer response times compared to less exhausted individuals. Additionally, it is expected that exhausted individuals will demonstrate a greater Stroop interference effect than less exhausted individuals on the classic Stroop task.

In the undertaking of this study, the following hypotheses will be tested:

1. Participants reporting high levels of burnout will perform worse on all of the Stroop blocks compared to participants reporting low levels.
2. Participants reporting high levels of anxiety and depression will perform worse on all of the Stroop blocks compared to participants reporting low levels.
3. Participants reporting high levels of perceived stress will perform worse on all Stroop blocks compared to participants reporting low levels.
Method

Participants

In total, 56 participants were recruited through ads on Facebook, flyers distributed within Umeå University, and by personal contact. Data were first scanned for outliers using IBM SPSS Statistics 24, four participants were excluded from further analyses as a result of their total score on the Stroop blocks being identified as extreme values (i.e., scores that were more than 2 standard deviations from the mean performance of the sample). As suggested by Field (2013), extreme values are not considered representative of the sample, and should therefore be excluded. Of the 52 remaining participants, 49 were students enrolled in full-time higher education and three were employed working full-time. Thirty self-identified as female and 22 as male; the age of the participants ranged from 18-52 years (M=25.3, SD=8.4). Forty-seven participants had not received any treatment for mental illness in the last six months; five individuals had received mental health treatment, of which three for depression, one for stress-related mental illness, and one unspecified treatment. The inclusion criteria to participate in the study required participants to be 18 years or older, to have not been diagnosed with reduced color sensitivity and speak Swedish fluently. Participants were offered coffee and cookies as compensation for their time involved in the study.

Instruments and materials

Burnout. The Shirom-Melamed Burnout Questionnaire (SMBQ; Grossi et al., 2003; Melamed et al., 1992) was used to measure participants’ perceived level of burnout. The self-assessment questionnaire contains 22 items, divided into four subscales: Physical fatigue (8 items), Cognitive weariness (6 items), Tension (4 items), and Listlessness (4 items). Each item is rated using a seven-point scale (1=Almost never to 7=Almost always; Lundgren-Nilsson, Jonsdottir, Pallant, & Ahlborg, 2012), the total score is comprised of the mean scores of all items (Perski & Grossi, 2018). The SMBQ correlates with the subscale Emotional exhaustion in Maslach Burnout Inventory (r = .77) and Pines Burnout Measure (r = .87; Grossi et al., 2003). The SMBQ also correlates with the clinical diagnosis ED (Glise et al., 2012). The cut off scores used in the present study for SMBQ are based on previous research (Grossi et al., 2003) where ≤2.75 was considered low burnout, and ≥3.75 was considered high burnout. According to the SMBQ manual (Perski & Grossi, 2018) ≥4.75 is considered pathological. Cronbach’s alpha for the present study were .92.

Perceived Stress. The Perceived Stress Scale (PSS-10) was used to measure participants’ individual assessments of their perceived levels of stress. The Swedish translation of the test used in the present study was validated by Nordin and Nordin (2013). The questionnaire contains 10 items and uses a four-point scale (0=Never to 4=Almost always) and the total scale ranges from 0 to 40 where a high score indicates high levels of perceived stress. The Cronbach’s alpha in previous research has been reported as .84 (Nordin & Nordin, 2013), in the present study it was .87.

Anxiety and Depression. The Hospital Anxiety and Depression Scale (HADS; Lisspers, Nygren, & Soderman, 1997; Zigmond & Snaith, 1983) was used to measure participants’ perceived levels of anxiety and depression. The HADS was first developed for general medical clinics and is now widely used for both clinical and research purposes (Crawford, Crombie, & Taylor, 2001; Herrmann, 1997). The HADS is a 14-item self-assessment questionnaire that is divided into two subscales with 7 items each, HADS-A for the Anxiety scale and HADS-D for the Depression scale. The questions are rated on a likert scale
from 0 to 3. Previous studies note the Cronbach’s alpha varied from .68 on HADS-A to .93 on HADS-D (Bjelland, Dahl, Tangen Haug, & Neckelmann, 2002). Although there is no consensus on cut-off scores to be used for HADS-A and HADS-D (Herrmann, 1997), the authors of the test recommend that a raw score between 8 to 10 might be considered as mild cases of anxiety or depression (Snaith & Zigmond, 1994, referred in Crawford et al., 2001). HADS has been shown to correlate with Becks Depression Inventory ($r = .73$), and the State-Trait Anxiety Inventory ($r = .68$ to $r = .71$; Lisspers et al., 1997). Cronbach’s alphas for the present study were .79 on HADS-A and .80 on HADS-D.

**Cognitive Performance.** The Modified Stroop Task was developed in the present study using the program ‘Psychopy’, and consists of five blocks with 100 stimuli in each block. During the duration of the task, participants were instructed to identify the color of which the stimuli were presented and press the corresponding key as fast and correct as possible (i.e., arrow up on the keyboard was colored red, arrow down green, left yellow and right blue) in line with previous research (Blanchette & Richards, 2012; Salo, Robertson, & Henik, 2001). There was a practice trial consisting of 20 stimuli representing each of the five experimental blocks (for a summary of the blocks, see Table 2). The first block consisted of color words, where the words and the ink-colors were congruent. The second block consisted of 17 neutral words, with a mean number of 6.06 characters. The words comprising this block were retrieved from a previous study in which they were deemed to be neutral in emotional valance; these words were specifically used as a neutral block in comparison with blocks of emotional words (Frings et al., 2010). Words such as “silence” and “chair” were presented randomly 100 times (see Table 3). The third block consisted of a Color-naming task where the stimuli “XXXXX” was presented. The purpose with the Color-naming block was to use it in analysis as a control for a Stroop effect compared to other blocks, as it has been used in other studies (MacLeod, 1991; Salo, Robertson, & Henik, 2001; Stroop, 1935). The fourth block consisted of 17 Exhaustion words, with a mean number of 7.35 characters, retrieved from the items in the SMBQ. The SMBQ is a validated questionnaire that aims to measure levels of burnout; therefore, for the purposes of the present study identifying the key words from the items comprising the scale was deemed appropriate. This was done by examining every item in the questionnaire and identifying key-words; for example, from the item worded, “I feel tired” the word “tired” was selected. Words such as “lazy” and “tense” were presented randomly 100 times in this block (for a complete list of the Exhaustion and Neutral words, see Table 3). The fifth block consisted of incongruent color words (the color word and the color it was printed in was incongruent). The order of the blocks was randomized and counterbalanced so that 50% of the participants performed the task in the order: Congruent, Neutral, Color-naming, Exhaustion and Incongruent (version A), while the other 50% performed the task in the opposite order (version B). These versions were administered so that the presentation of the test protocol would be counterbalanced, based on similar research (Blanchette & Richards, 2012). The two blocks with Neutral and Exhaustion words were based on the theory of attentional bias to threat words (Frings et al., 2010), and were compared to each other in subsequent analyses. The Congruent, Incongruent, and the Color-naming block were constructed based on the original Stroop task (MacLeod, 1991; Stroop, 1935). The whole task was administered on a MacBook Pro and an Asus K56CM. Version A and B were administered an equal number of times on both computers.
Table 2

Example of items within the blocks in the Modified Stroop Task (printed color of the item).

<table>
<thead>
<tr>
<th>Congruent</th>
<th>Neutral</th>
<th>Color-naming</th>
<th>Exhaustion</th>
<th>Incongruent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue (blue)</td>
<td>Paper (blue)</td>
<td>XXXXX (green)</td>
<td>Tense (red)</td>
<td>Blue (red)</td>
</tr>
<tr>
<td>Red (red)</td>
<td>Format (green)</td>
<td>XXXXX (red)</td>
<td>Passive (blue)</td>
<td>Red (blue)</td>
</tr>
<tr>
<td>Green (green)</td>
<td>Corner (yellow)</td>
<td>XXXXX (blue)</td>
<td>Drowsy (green)</td>
<td>Green (yellow)</td>
</tr>
<tr>
<td>Yellow (yellow)</td>
<td>Month (red)</td>
<td>XXXXX (yellow)</td>
<td>Tired (yellow)</td>
<td>Yellow (green)</td>
</tr>
</tbody>
</table>

Table 3

Neutral and Exhaustion words used in the Modified Stroop Task. English translations and the Swedish words used as items in the administered version.

<table>
<thead>
<tr>
<th>Neutral English</th>
<th>Neutral Swedish</th>
<th>Exhaustion English</th>
<th>Exhaustion Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silence</td>
<td>Tystnad</td>
<td>Tired</td>
<td>Trött</td>
</tr>
<tr>
<td>Period</td>
<td>Period</td>
<td>Dull</td>
<td>Matt</td>
</tr>
<tr>
<td>Winter</td>
<td>Vinter</td>
<td>Exhausted</td>
<td>Utmattad</td>
</tr>
<tr>
<td>Chair</td>
<td>Stol</td>
<td>Lack of energy</td>
<td>Energilös</td>
</tr>
<tr>
<td>Square</td>
<td>Fyrkant</td>
<td>Drained</td>
<td>Utförd</td>
</tr>
<tr>
<td>Paper</td>
<td>Papper</td>
<td>Lazy</td>
<td>Slö</td>
</tr>
<tr>
<td>Format</td>
<td>Format</td>
<td>Burned out</td>
<td>Utbränd</td>
</tr>
<tr>
<td>Corner</td>
<td>Hörn</td>
<td>Decrepit</td>
<td>Orkeslös</td>
</tr>
<tr>
<td>Structure</td>
<td>Struktur</td>
<td>Passive</td>
<td>Passiv</td>
</tr>
<tr>
<td>Bottle</td>
<td>Flaska</td>
<td>Drowsy</td>
<td>Dåsig</td>
</tr>
<tr>
<td>Month</td>
<td>Månad</td>
<td>Tense</td>
<td>Spänd</td>
</tr>
<tr>
<td>Board</td>
<td>Styrelse</td>
<td>Restless</td>
<td>Rastlös</td>
</tr>
<tr>
<td>Wool</td>
<td>Bomull</td>
<td>Tensions</td>
<td>Spännningar</td>
</tr>
<tr>
<td>Era</td>
<td>Epok</td>
<td>Poor concentration</td>
<td>Okoncentrerad</td>
</tr>
<tr>
<td>Angle</td>
<td>Vinkel</td>
<td>Slow thinking</td>
<td>Trögtänkt</td>
</tr>
<tr>
<td>Concept</td>
<td>Koncept</td>
<td>Complicated</td>
<td>Komplicerat</td>
</tr>
<tr>
<td>Example</td>
<td>Exempel</td>
<td>Fragmented</td>
<td>Splittrad</td>
</tr>
</tbody>
</table>

Procedure

The test protocol was first pilot tested with a sample size of 5 participants (10% of the desired total sample) and included participants that represented the anticipated sample in the main study. The test battery was examined, and small adjustments were made based on the feedback from the pilot testing. For example, information regarding how many blocks
remaining during the test was added, as well as an instruction to remember to stay focused during the whole task.

The test session lasted approximately 20 minutes and was administered on a computer in a quiet room. The participants started by reading and completing a consent form. They also provided demographic information consisting of age, gender, occupation, level of education, if they have received any treatment for mental illness in the last six months, and if they have any diagnoses that can affect their cognition. Participants then completed the Modified Stroop Task and were instructed to carry out the task as fast and as accurately as possible. Lastly, they completed the self-assessment questionnaires (i.e., SMBQ, HADS, PSS-10). Upon completion of the testing, participants were debriefed regarding the nature of the research and were then offered coffee and cookies for their participation and to allow time for participants to ask questions.

Data analysis

Data were analysed using SPSS. Raw scores reflecting response times and errors on the Stroop task were automatically entered into an MS Excel file generated by the Psychopy programme and a performance score for each block was calculated using the format: total time + ((total time/number of stimuli) × number of errors) (Gardner et al., 1959, referred in Scarpina & Tagini, 2017). Hence, a high score on the Stroop task represents worse performance.

In order to compare performance between version A and version B in the Modified Stroop Task, a randomization check was carried out using a MANOVA. Mean values on the Stroop blocks were compared using one-way repeated measures ANOVA. Median splits were also made based on the scores on the SMBQ, HADS and PSS-10 to enable between group comparisons on dependent variables using MANOVAs. Descriptive statistics were used to describe demographic variables among the participants.

Ethical considerations

Prior to participating, all participants provided informed verbal and written consent and were reminded that their participation was voluntary. They were also informed that they had the right to discontinue the test at any time without being required to provide a reason. The data were coded to ensure the participants’ anonymity.

Results

Randomization check

The Modified Stroop Task was administered in two versions, in which the blocks were presented in different orders (version A and version B detailed in the Method section), a MANOVA was undertaken to test if performance within the Stroop blocks differed between the two versions. There were no significant differences in performance on any of the blocks between version A and version B. Therefore, the following analyses represent all combined data from both versions reflecting performance within the blocks.

Main analyses

To examine the research question regarding potential significant differences in performance across the Stroop blocks, a one-way repeated measures ANOVA was conducted. When the assumption of sphericity was violated, a Greenhouse–Geisser adjustment was
applied to the degrees of freedom. There was a significant difference in performance across the Stroop blocks, $F(1.982, 101.062) = 9.95, p < .001$. Bonferroni follow-up tests revealed that performance on the Incongruent block was significantly worse than performance on the Neutral and the Color-naming block $p < .05$, additionally it was approaching a significant difference with the Congruent block ($p = .061$). Performance on the Exhaustion block was significantly worse than performance on the Color-naming block, $p < .05$. Additionally, performance on the Neutral block was significantly worse than the Color-naming block, $p < .05$ (see Table 4).

To examine whether there were significant differences between participants scoring high or low on the SMBQ on Stroop performance across the blocks, median-splits on SMBQ scores were undertaken ($Md=63.5, SD=17.98$) identifying a low-exhaustion group ($n=26$) and a high-exhaustion group ($n=26$). A MANOVA comparing Stroop performance between high and low exhaustion groups revealed no significant differences in performance on any of the blocks between these groups, Pillai’s Trace $= .184, F(1, 50) = 2.073, p = .086$.

To examine the clinical designation of exhaustion, two groups were also identified based on the SMBQ-global cut off value for exhaustion ($\geq 3.75$) as well as the cut off value for healthy individuals ($\leq 2.75$). Participants were divided into a low-exhaustion group ($n=17$) and a high-exhaustion group ($n=11$). There were no significant differences in performance on any of the blocks between these groups, Pillai’s Trace $= .190, F(1, 26) = 1.031, p = .424$. However, there was a general tendency in all blocks with the high-exhaustion group performing less well than the low-exhaustion group.

Table 4

| Mean values (and standard deviations) of performance in the Modified Stroop Task. |
|---------------------------------|-----------------|-----------------|
| Stroop block                    | Overall         | Low exhaustion (≤2.75) | High exhaustion (≥3.75) |
| Color-naming block              | 65.51 (6.7)     | 66.81 (8.59)       | 68.08 (3.85)           |
| Congruent block                 | 67.98 (10.38)   | 67.99 (11.91)      | 72.62 (7.04)           |
| Neutral block                   | 68.18 (7.77)    | 67.99 (9.34)       | 72.87 (6.60)           |
| Exhaustion block                | 69.10 (7.36)    | 69.55 (8.47)       | 72.47 (7.44)           |
| Incongruent block               | 73.13 (13.35)   | 75.39 (19.18)      | 76.04 (7.74)           |

To investigate whether there were significant differences in performance across the Stroop blocks between high and low scoring participants on the HADS, median-splits were undertaken in order to compare participants with low and high scores on the HADS-A ($Md=5.00, SD=3.38$) and HADS-D ($Md=2.00, SD= 3.15$). Participants were first divided into a low-anxiety group ($n=20$) and a high-anxiety group ($n=22$). A MANOVA revealed there were no significant differences in performance on any of the blocks between these groups, Pillai’s Trace $= .129, F(1, 40) = 1.064, p = .396$. Participants were also divided into a low-depression group ($n=14$) and a high-depression group ($n=25$). A MANOVA revealed there were no significant differences in performance on any of the blocks between these groups, Pillai’s Trace $= .114, F(1, 37) = .851, p = .524$.

To examine the potential influence of clinical levels of anxiety and depression, two groups were identified based on the HADS cut off value for anxiety (i.e., 8) as well as the cut off value for depression (i.e., 8). Participants were divided into a low-anxiety group ($n=41$) and a high-anxiety group ($n=11$). A MANOVA revealed there were no significant differences in performance on any of the blocks between these groups, Pillai’s Trace $= .063, F(1, 50) = .620, p = .424$.
p = .686. Participants were divided into a low-depression group (n=14) and a high-depression group (n=25). A MANOVA revealed there were no significant differences in performance on any of the blocks between these groups, Pillai’s Trace = .030, F(1, 50) = .288, p = .917.

To investigate whether there were significant differences in Stroop performance between high and low scoring participants on the PSS-10, a median-split (Md=13.00, SD=5.87) was undertaken; participants were divided into a low-stress group (n=22) and a high-stress group (n=24). A MANOVA revealed there were no significant differences in performance on any of the blocks between these groups, Pillai’s Trace = .160, F(1, 44) = 1.529, p = .203.

Discussion

The aim of the present study was to explore potential differences in performance on a Modified Stroop Task comprised of Exhaustion-related words. In particular, participants’ levels of burnout, anxiety, depression, and stress were compared with performance. The findings arising from analyses highlight significant differences in performance between the Stroop blocks. More specifically, performance on the Incongruent block was significantly worse than on the Neutral and Color-naming block. Furthermore, performance on the Exhaustion and Neutral block were significantly worse than on the Color-naming block. There were no significant differences in performance on the Stroop tasks and participants’ scores on measures of burnout, anxiety and depression, or stress; although, results were approaching significance in comparison of high scores on burnout and worse Stroop performance.

The difference in performance between the Stroop blocks indicates that a Stroop effect occurred, in accordance with previous research on the Stroop effect (MacLeod, 1991; Stroop, 1935). In consideration that both the Incongruent and Exhaustion blocks induced the largest Stroop effect, compared to the Color-naming block, it can be deduced that the test manages to capture a Stroop interference effect within the original Stroop blocks as well as in the Exhaustion and Neutral blocks. There was also a general tendency that participants above the cut-off score on the burnout scale performed worse on all Stroop blocks, this tendency was not found when participants were grouped based on median-splits of burnout. This could indicate that the Modified Stroop Task may be more sensitive to differentiate highly exhausted individuals.

In attempting to explain why the results in the present study did not reveal a statistically significant difference between high and low exhausted participants, and their performance on the Stroop task, previous studies demonstrating exhaustion’s influence on cognition and executive functions may be considered (Deligkaris et al., 2014; Diestel et al., 2013; Ellbin et al., 2017). Although a trend approaching significance was observed in the present study, one possible explanation for it not being significant may relate to previous research using groups of clinically burned out participants, when measuring performance on cognitive tasks (Deligkaris et al., 2014; Diestel et al., 2013; Ellbin et al., 2017). Participants in the present study were mostly students that did not meet the criteria for ED. The cut-off score used to group the participants into a high and low exhaustion group was the SMBQ-Global score ≥3.75 for high burnout and the non-pathological healthy score was ≤2.75 (i.e., cut off-scores used by Grossi, 2003). The SMBQ-Global cut-off for high exhaustion was used in the present study, as there were too few participants reporting pathological cut-off value (i.e., ≥4.47 proposed by Perski & Grossi, 2018). Therefore, the highly exhausted group in the present study may not have suffered the cognitive impairments to the same degree that the pathological exhausted participants in previous studies (Deligkaris et al., 2014; Diestel et al., 2013; Ellbin et al., 2017). Future studies should consider this limitation when determining sample characteristics and
recruiting participants. The use of a clinical based sample diagnosed with ED could strengthen the interpretation of results as there would be greater homogeneity of the sample.

The Modified Stroop Task was intended to measure cognitive functions, similar to the original Stroop (MacLeod, 1991; Stroop, 1935) as well as attentional bias captured by the Emotional Stroop Tasks (Cisler & Koster, 2010; Phaf & Kan, 2007). However, the Modified Stroop Tasks’ result did not demonstrate similar findings as other Emotional Stroop Tasks (Cisler & Koster, 2010; Phaf & Kan, 2007), therefore it is not possible to be completely confident that it measured the intended attentional bias towards threat-words. Additionally, although the original Stroop task is well validated, the Modified Stroop Task is a modification inspired by the original test and requires further validation. To ensure the validity of the Modified Stroop Task, future researchers could compare the result to a validated version of the Stroop task (e.g., Golden & Freshwater, 2002). If comparable test results correlate with each other the validity of the Modified Stroop Task would be enhanced as it demonstrates that it measures the cognitive functions it purports to; that is, the Stroop Effect as well as attentional bias.

However, the results of the present study show a significant difference between the Exhaustion block and Color-naming block as well as between the Neutral and Color-naming block, this may be due to the process that reading words can create longer response times (i.e., the Stroop interference effect), as it has been reported in previous studies (MacLeod, 1991; Stroop, 1935). However, as there is no significant difference between the Exhaustion block and the Neutral block, it indicates that attentional bias towards exhaustion-related words are not larger than towards neutral words, as it has been observed in other Emotional Stroop Tasks (Cisler & Koster, 2010; Phaf & Kan, 2007). Thus, the difference between the Stroop blocks seems to be dependent on the reading process, rather than on attentional bias towards threat-words. In contrast, another result pointing to the fact that the Modified Stroop Task may be able to capture an attentional bias, is that performance on the Incongruent block is significantly worse than on the Neutral block, but not than on the Exhaustion block. This could indicate that there is an influential factor present in the Exhaustion block that augments response times, and this factor is not present in the Neutral block. Additionally, the Stroop effect (e.g., the difference between the Color-naming block and the Incongruent block) that emerged in the present study can imply that the construction of the Modified Stroop Task is an accurate modification of the original Stroop test.

Furthermore, there appears to be no previous research where an Emotional Stroop Task has been modified for exhaustion; since the present study found no significant differences between high and low exhausted people, attentional bias cannot be identified as a characteristic of those with exhaustion. Although, attentional bias appears to be present in all anxiety disorders (Cisler & Koster, 2010), and since exhaustion has been identified to relate with anxiety (Ding et al., 2014; Zhou et al., 2016), it was assumed that attentional bias would be present among those with high exhaustion. The reason that this could not be demonstrated in the present study may be due to the sample not being comprised of participants with pathological exhaustion. Therefore, it is not a definitive conclusion that attentional bias does not occur among exhausted individuals, and future studies should therefore contemplate ways of further evaluating this. This could be done by carrying out more studies where Emotional Stroop Tasks are modified for exhaustion, in which the limitations of this study are considered.

Even though the findings of the present study offer insight into exhaustion and Stroop performance, the construction of the Modified Stroop Task also requires discussion. In particular, the development of the test protocol regarding the stimuli in the Modified Stroop Task warrants scrutiny. The Neutral words used were developed and validated in Germany (Frings et al., 2010) and may not be representative for a Swedish population, due to assumed
cultural differences as well as differences in connotation between the Swedish and the German translations of the words. Other researchers (e.g., Blanchette & Richards, 2012) used non-words to ensure the participants did not have any pre-existing associations with the words. In the present study, the decision to use Neutral words was based on meta-analyses of Emotional Stroop Tasks (Cisler & Koster, 2010; Phaf & Kan, 2007) and the process of producing non-words was considered a risk of reducing the credibility of the test. Further studies should consider executing a study where participants that match the intended sample in the Modified Stroop Task get to estimate their perception of words that are considered neutral. In so doing, researchers can be assured that words in the Neutral block are perceived as neutral and therefore reduce the risk of attentional bias towards the words in this block.

The words used in the Exhaustion block also requires discussion. The decision to use words retrieved from the SMBQ was considered appropriate as it has been validated for use in the measurement of burnout. A possible limitation with this method could be that SMBQ measures burnout rather than clinical ED; however, exhaustion has been identified as the core dimension of the burnout concept (Maslach et al., 2001). An alternative approach for future research could include the use words retrieved from the proposed diagnostic criteria for ED recommended by the Swedish National Board of Health and Welfare (2003). Consultation with practitioners in the community that undertake the evaluation of individuals for ED could improve both construct and ecological validity of the words used within the Exhaustion block.

Additionally, the length of the words (i.e., number of letters comprising the word) is also a factor that could have affected the results. The mean length of the Exhaustion words was slightly longer than the mean length on the Neutral words, this could have contributed to a small difference in performance between these blocks, with slower response times due reading of longer words. However, there was no significant difference between these blocks; therefore, it is difficult to determine whether the length of the words contributed to the observed results.

Another matter regarding the Modified Stroop Task, relates to the number of items presented. The decision to include the presentation of 100 stimuli in each block was partly based on the assumption that the task had to be long enough so that potential differences between the blocks could be observed. If a larger number of stimuli were used, there may have been a greater difference in performance between the Stroop blocks and between groups. On the other hand, if the number of stimuli was excessive, performance could be biased due to participant fatigue. This was taken under consideration when designing the Modified Stroop Task, by adding instructions throughout the task about the number of remaining blocks. Previous research varies in the number of items used in the Stroop Task (e.g., Blanchette & Richards, 2012; Salo et al., 2001), therefore no general standard for how many stimuli that should be included has been established.

A further possible improvement in future studies is regarding the blocks included in the Modified Stroop Task. The Congruent block is considered redundant, since there is a bigger Stroop interference between the Color-naming block and the Incongruent block in other studies (MacLeod, 1991; Stroop, 1935), as well as in the present study. An additional argument for excluding the block is that there were no significant differences between the Congruent block and the other blocks. Another consideration in future research is whether to include blocks based on the original Stroop task, researchers could contemplate having a narrower focus on attentional bias among exhausted individuals. This could be done by only including the Exhaustion and Neutral block.

The scoring method used for the Modified Stroop Task was calculated using a format presented by Scarpina and Tagini (2017). This format takes response time as well as number of errors into account, therefore it was considered to be an adequate method for calculating performance on the Stroop task. However, this format has been criticized, with the suggestion
that no available scoring format manages to include all the essential parts when calculating the Stroop score (Scarpina & Tagini, 2017). Future research may endeavor to use alternative scoring formats that weigh errors and response times differently in determining performance.

The test protocol also included self-assessment questionnaires; the decision to include questionnaires for anxiety, depression, stress, and burnout was based on research pointing towards these conditions being interrelated (Ahola et al., 2005; Ding et al., 2014; Hakanen & Schaufeli, 2012; Wang, 2005; Wiegner et al., 2015; Zhou et al., 2016). The fact that there were no significant differences associated with anxiety, depression, stress and burnout on Stroop performance, may be due to the construction of the Modified Stroop Task rather than the choice of questionnaires. However, the internal consistency on HADS-A (Bjelland et al., 2002) is considered inadequate (EFPA, 2013), neither are there any adequate norms and cut-off scores available for HADS. Therefore, future research may consider using a measurement of anxiety with greater reliability. One strength of the HADS is that it does not take long to complete, but this is not considered to compensate for these noted limitations.

The decision to use the SMBQ, was based on the fact that people with ED also scores high on burnout-scales (Glise et al., 2012; 2014). However, previous research presenting the Cronbach’s Alpha of the SMBQ is limited and constitutes a limitation in the reliability of the results regarding the SMBQ. A consideration is therefore whether SMBQ is the most valid measure of exhaustion. Future researchers should contemplate if there are benefits of using a questionnaire developed to measure exhaustion instead of burnout (e.g., Karolinska Exhaustion Disorder Scale; Besèr et al., 2013)

As there are overlapping symptoms within exhaustion and stress, the PSS-10 was included to ensure that the assessments on SMBQ and PSS-10 correlate, as has been observed in other studies (Nordin & Nordin, 2013). A correlation between these questionnaires would indicate that scorings are reliable. However, it may not be necessary to include PSS-10, since scores on these scales already have been shown to often correlate, and also measures similar conditions.

Another important consideration when developing the Modified Stroop Task was determining the operational definition of burnout used to formulate the components and words comprising the Modified Stroop Task. Theories on clinical burnout have emphasized various components related to the condition. The present study was conducted based on both Maslach’s (2001) and Melamed’s definitions of burnout. Although, one could argue that the definition by Melamed (1992) had the greatest impact on the development of the test protocol, since the stimuli in the Modified Stroop Task were created based on the SMBQ. Also, Melamed (1992) highlights cognitive weariness as a symptom of clinical burnout, and cognitive impairments emerges as a common symptom in patients with ED in recent studies (Diestel et al., 2013; Deligkaris et al., 2014; Ellbin et al., 2017). In light of both Maslach and Melamed’s definitions including emotional exhaustion, both theoretical approaches of burnout, as well as research about cognitive impairments among exhausted individuals (Diestel et al., 2013; Deligkaris et al., 2014; Ellbin et al., 2017), influenced the development of the Modified Stroop Task. There was a general tendency in the present study for highly exhausted participants to perform worse and demonstrate a greater Stroop interference effect; this could suggest that the theoretical background of the development of the test was sound.

Furthermore, research on stress and exhaustion indicates that these conditions often overlap with depression and anxiety (Ahola et al., 2005; Ding et al., 2014; Haakanen & Schaufeli, 2012; Wang, 2005; Wiegner et al., 2015; Zhou et al., 2016). Alternatively, some researchers posit that these conditions share similar symptomatic characteristics, but argue that they should be considered as distinct states (Iacovides et al., 2002). Based on this, the present study hypothesized that the Modified Stroop Task might also be sensitive in capturing
depression, anxiety, and perceived stress. However, as the results indicate no differences in performance between high and low depression, anxiety or stress, the Modified Stroop Task does not seem to differentiate these groups. Additionally, highly exhausted participants performed worse than low exhausted participants on all Stroop blocks, and this trend was not observed between high and low scores on any other measurement; therefore, this may indicate the Modified Stroop Task is more sensitive in capturing exhaustion than any of the other related states. This, combined with the fact that the trend was not significant, indicates the need for further research on assessment of exhaustion that can separate these states from each other.

A further consideration of the present study is regarding the sample; it consisted of 52 participants included in the analysis, ranging from 18 to 52 years, with the majority being students \((n=49)\). This is considered a small sample, since guidelines on evaluation of psychological tests propose that less than 200 participants are considered inadequate (EFPA, 2013). Additionally, participants were recruited randomly at a university and constituted a convenience sample. This is considered a limitation, since this group is not representative of a larger Swedish population. Therefore, it is not reasonable to propose that future studies would observe the same results if carried out with a wider sample. ED is a common condition among the working population in Sweden (Försäkringskassan, 2015) with research and theories developed around the links between exhaustion and job performance (Bakker & Demerouti, 2017; Bakker et al., 2008; Taris, 2006), a case could be made that a sample of a working population may enhance the ecological validity of the present study. Moreover, the effect of aging is a factor that has been consistently highlighted in research examining performance on cognitive tests (e.g., the Stroop test; Milham et al., 2002; Verhaeghen & Cerella, 2002). Age is an important factor to consider when undertaking future studies to extend the present study. Thus, in consideration of the sample size and its composition, it warrants the statement that the present study may serve as a pilot study for use in future research.

In summary, the present study serves as an innovative attempt to develop an instrument aiming to identify exhausted individuals. Further development of the Modified Stroop Task may lead to its use in screening individuals seeking help for stress-related mental illness (e.g., ED). With further validation it may offer researchers and clinicians a means to screen for ED that negates the limitations of self-report measures. In consideration that the symptoms of ED overlap with similar conditions, it warrants the statement that there is a need for methods to differentiate between conditions to support the prescription of optimal treatments. A fully validated version of the Modified Stroop Task could serve to evaluate treatment outcome for patients with ED, especially regarding the cognitive impairments that patients often report (Jonsdottir et al., 2013). The preliminary findings of the cognitive performance measure developed in the present study offer a foundation for future research that may lead to new methods in identifying and evaluating the treatment of ED and reducing its impact on the wider society of Sweden.
Reference list


EFPA. (2013). EFPA review model for the description and evaluation of psychological and educational tests: Test review form and notes for reviewers (4.2.6).


