Measuring Vulnerability to Depression:
The Serbian Scrambled Sentences Test – SSST

Zdenka Novović, Ljiljana Mihić, Mikloš Biro, and Snežana Tovilović
Faculty of Philosophy, Novi Sad, University of Novi Sad, Serbia

The goal of this study was to establish whether the SSST, a Serbian language scrambled sentences instrument, is a reliable measure of depressive cognitive bias, and whether it captures the suppression tendency as participants exert the additional cognitive effort of memorizing a six-digit number while completing the task. The sample consisted of 1071 students, randomly assigned into two groups. They completed the SSST divided into two blocks of 28 sentences, together with additional cognitive task during either the first or second block, and after that a number of instruments to establish validity of the SSST.

The test was shown to be a reliable instrument of depressive cognitive bias. As a measure of suppression the SSST performed partly as expected, only when load was applied in the second half of the test, and fatigue and cognitive effort enhanced suppression. The advantages of the test versus self-description measures were discussed.

Keywords: psychometric characteristics, cognitive bias, suppression, vulnerability, dysphoria

The concept of vulnerability to depressive disorders (whether as inherited predisposition or acquired) is closely connected to the idea of cognitive schemas. According to Beck (1967), cognitive schemas are enduring memory structures which are used in organising new information (Clark, Beck, & Alford, 1999). In persons vulnerable to depression, schemas contain deeply embedded absolutistic beliefs of worthlessness, incompetence, and unlovability (Dozois & Beck, 2008). As the basic and deepest level of cognitive organisation, the cognitive schema influences the next, intermedial level above it, dysfunctional assumptions and rules and cognitive biases (Dozois & Beck, 2012). Beck’s hypothesis holds that the latent schema is dormant until activated by a congruent event, i.e. a stressor

Corresponding author: zdenov@ff.uns.ac.rs

The study is part of the project of Ministry of Education, Science and Technological Development, “Hereditary, environmental, and psychological factors of mental health”, no 179006. The standardization of SSST is supported by a Provincial Secretariat for Science and Technological Development as a part of project no. 114–451–1647/2011–01. The authors would like to thank Veljko Jovanović for his help in construction of scrambled sentences, and Vera Maletić for help in editing the answer key for the test.
resembling the event responsible for the disfunctional formation of that schema in the past. As the depressive symptoms abate, the schema once again becomes inactive (Beck, Rush, Shaw, & Emery, 1979; Sacco & Beck, 1995).

Until the 1990s, the most widely used instrument for measuring deeper cognitive structures was the DAS or Dysfunctional Attitude Scale (Weissman & Beck, 1978). However, studies using this instrument to ascertain the stability of cognitive schemas failed to show that the latent depressive schema, as measured by the DAS, was in fact a lasting trait of vulnerable individuals (Dykman, 1997; Eaves & Rush, 1984). In the 1990s, some researchers succeeded in activating depressive cognitive schemas in vulnerable individuals by inducing affect in remitted depressed patients (Ingram, Miranda, & Segal, 1998). However, the results of such studies were inconclusive (Ingram et al., 1998).

Since then, a substantive body of research has been conducted suggesting a role of cognitive reactivity in schema activation in remitted depressives. Namely, even mild mood fluctuations can trigger maladaptive cognitions in formerly depressed individuals (Ingram, Atchley, & Segal, 2011). However, another important question has still remained unanswered – how does the schema become deactivated following depression (Wenzlaf, Rude, & West, 2002).

To explain schema deactivation, Wenzlaff and colleagues have pointed out that individuals who had experienced a depressive episode tend to suppress negative thoughts, i.e. to consciously and deliberately put unpleasant thoughts out of mind, resisting depressive affect (Wenzlaff & Bates, 1998; Wenzlaff & Eisenberg, 2001; Wenzlaff, Rude, Taylor, Stultz, & Sweatt, 2001; Wenzlaf et al., 2002). Several studies have also suggested that suppression merely masks vulnerability, and that vulnerable persons will return to negative thinking whenever their mental control is undermind by other cognitive demands (Wenzlaff & Bates, 1998; Wenzlaff & Eisenberg, 2001). The authors thus alter the meaning of latent cognitive schema. They consider that cognitive schemas in formerly depressed persons remain active, and are latent merely in the sense that they temporarily lack conscious focus (Najmi & Wegner, 2009).

Research has shown the inadequacy of suppression as an avoidance mechanism, for intrusive thoughts return with often greater frequency despite attempts to suppress them (Najmi & Wegner, 2009). Wegner explained this rebound effect by Ironic Process Theory, postulating that suppression relies on two mechanisms: one conscious and the other unconscious (Wegner, 1994; Wegner & Erber, 1992). The conscious operating process is used in searching for a thought to replace an unpleasant idea; the unconscious monitoring process requires minimal cognitive effort and is always on the lookout for suppressed thoughts which may enter the consciousness, triggering the conscious operating process upon detection. Successful suppression requires the interaction of both processes. Ironically, it is the unceasing vigilance of the monitoring system itself which prevents unwanted thoughts from ever fully subsiding.

Those vulnerable to depression are poorly equipped to suppress, for several reasons. Stress, affect and cognitive demands overload executive
memory capacity, in turn impairing the conscious operating process (Brewin & Smart, 2005). Also, during distraction from negative thoughts, other unpleasant content tends to attract attention as it is connected to suppressed thoughts by means of affective valence (Bower, 1987; Wenzlaff, Wegner, & Roper, 1988); hence, further attempts at suppression only increase the prevalence of negative thoughts (Wenzlaff, 2005).

In order to disable mental control and tap into more automatic cognitive processes in depression, Wenzlaff and colleagues innovated previous laboratory procedures used in research on processing biases. Namely, participants are given tasks with ambiguities and must perform them while also handling other cognitive demands such as memorising strings of numbers. The true purpose of the research is not revealed, allowing vulnerable persons to show depressive bias through their responses. The additional cognitive load prevents suppression, hindering the conscious operating system in its attempts to distract negative thoughts (Phillips, Hine, & Thorsteinsson, 2010). The tendency to suppress is indicated by the difference between results obtained with and without the additional cognitive load.

The most widely used measure of negative interpretation bias and suppression in depression is the SST or Scrambled Sentence Test (Wenzlaff & Bates, 1998). In their first study using the SST to test Ironic Process Theory, Wenzlaff and Bates (1998) assigned sentence-formation tasks to vulnerable, non-vulnerable, and depressive participants. A number of participants in all three groups was also given additional cognitive load. The authors confirmed that vulnerable participants perform like non-vulnerable participants when solving sentences without cognitive load, but like depressive participants when additional cognitive effort is required. These results support the hypothesis that cognitive load impairs mental control mechanisms in vulnerable persons, thereby making it possible to measure suppressed depressive content. Watkins and Moulds (2007) replicated the results in a clinical population, with formerly depressed patients performing on the SST under cognitive load condition like currently depressed participants, rather than like the general population.

Following Wenzlaff and Bates, other researchers also used scrambled sentences. In two experiments, Rude and colleagues found that negative interpretation bias, shown during the cognitive-load condition on the SST, predicted future depressive symptoms and the diagnosis of major depression, particularly in combination with high suppression tendency. They further demonstrated that there were gender differences in exhibition of depressive cognitive bias i.e., better prediction of depression in men was obtained by a cognitive load procedure compared to women (Rude, Wenzlaff, Gibbs, Vane, & Whitney, 2002; Rude, Valdez, Odom, & Ebrahimi, 2003). In a recent study, Rude and associates compared the predictive power of the SST and of the DAS regarding subsequent symptoms of depression, finding that both instruments had significant independent predictive power (Rude, Durham-Fowler, Baum, Rooney, & Maestas, 2010). The DAS and the Serbian version of the SST were also compared by Tintarović and associates regarding dysphoria symptoms
in non-clinical populations, with similar results (Tintarović, Novović i Mihić, 2012). Several studies have used the SST to measure the effect of cognitive behavioral modification targeting interpretation. Studies demonstrated that the SST was sensitive enough to capture change in cognitions following therapy (Blackwell & Holmes, 2010; Holmes, Lang, & Shah, 2009; Holmes, Mathews, Dalgleish, & Mackintosh, 2006).

The goal of this study was to test psychometric characteristics of the Serbian version of the SST, and to show its performance in various sub-samples and under different conditions. Regarding its validity, it was examined whether the SSST acts as: a) a test of negative bias, and b) as a test of suppression. To establish its validity as a test of negative bias, one must show that the formation of depressive sentences is related to a greater extent to dysphoria and vulnerability to depression than to other affective states; to establish that the test prevents suppression of depressive thoughts, one must compare the results obtained with and without additional cognitive load. Another aim was to examine gender differences in our sample, as demonstrated in other studies (Rude et al., 2002). If found in our study, such results would lead to cross-cultural generalizability of the finding of greater suppression tendencies in men than women in western cultures. An ultimate aim of the study was to establish optimal procedures for administering the test and to provide reference values for future use.

**Method**

*Participants.* The sample consisted of 1071 participants: first- and second-year students of several faculties\(^1\) of the University of Novi Sad, 38% male. The average age was 19.59 (SD = 1.24).

*Instruments.* The Serbian Scrambled Sentences Test (SSST) was created in the Department of Psychology and modelled after the SST (Wenzlaff & Bates, 1998). A team of researchers formulated five-word sentences with depressive content using Wenzlaff’s original sentences and items containing the thoughts of depressive participants. These five words, plus a sixth word enabling the creation of a positive statement, were assigned numbers and randomly scrambled. An initial 60 tasks were created, with 56 retained after the pilot study\(^2\). Participants were informed that the test was a measure of verbal ability, and were asked to use 5 of the 6 words to create a meaningful declarative sentence as quickly as possible. Participants answered by writing the numbers 1 to 5 above the words they chose to use, indicating the desired order. Here is a sample task with a possible solution in the original language:

\[
\begin{array}{ccccc}
2 & 1 & 3 & 4 & 5 \\
MI & BUDUĆNOST & SVETLA & IZGLEDA & VEOMA & MRAČNA
\end{array}
\]

The solution shown above can be translated roughly as ‘my future looks very dark’, and would be rated as depressive, whereas the number 5 above the word *svetla* would yield ‘my future looks very bright’ and be rated as non-depressive. The number of depressive sentences was divided by the number of all grammatically correct sentences (according to the key made by a professional language editor), yielding the proportion of depressive sentences.

---

\(^1\) Arts and Humanities, Social Sciences, Natural Sciences, Engineering, and Medical Sciences.

\(^2\) Four items were excluded because of their ambiguous connotations. For example, a sentence “I worry about my future” in Serbian language can have both positive and negative meanings.
sentences to the total of grammatically correct sentences. The sentences were randomly split into two blocks of 28 sentences each, in order to administer half of the test with the added cognitive load of simultaneously memorizing a six-digit number. Additionally, the order of sentences within each block was randomized. Participants wrote the number after all sentences in the block. The cognitive load was administered in the counterbalanced order across the two groups of participants. The time limit for solving each block was 3.5 minutes, as recommended by American researchers (Rude et al., 2002). In previous SST studies, failure to reproduce the number did not eliminate participants, as the authors considered it to have no bearing on the results (Rude et al., 2002; Rude et al., 2003; Wenzlaff & Bates, 1998). However, Gilbert and Hixon (1991) noted that a failure to reproduce the digit might mean that participants did not comply with the procedure or that the task was highly effective in depleting cognitive resources. Hence, their recommendation was to exclude participants who remember incorrectly at least half of the numbers. Following this recommendation, we chose to eliminate participants who were unsuccessful in repeating at least 4 digits in any order. Of the 1162 participants, 86% repeated the number without errors, while 92.5% (1071) repeated at least 4 digits in any order.

Depression, Anxiety and Stress Scale (DASS; Lovibond & Lovibond, 1995) contains 21 items, 7 for each of the 3 dimensions measured. The Serbian translation of the scale demonstrated reliability and discriminant validity (Jovanović, Žuljević i Brdarić, 2011). Subscale reliability in our sample ranged from 0.75 (anxiety scale) to 0.78 (depression scale).

Psychological Distance Scaling Task (PDST; Dozois & Dobson, 2001a) is designed to assess the content organisation of cognitive schemas (Dozois & Dobson, 2001a). The instrument is based on the assumption that negative contents are more consolidated (i.e., have smaller interstimulus distances) than positive within a depressive cognitive schema (Dozois, 2007). The test is administered by computer, and consists of 80 adjectives describing personality traits which the participants rate regarding self-relevance and valence within a coordinate system. Along the x-axis, the degree of self-relevance was rated from very much like me to not at all like me; along the y-axis, valence was rated from positive to negative. Adjectives are pre-classified as positive (’intelligent’) or negative (’envious’), and in terms of social relations (’sociable’) or achievement (’successful’). Four scores are calculated: average interstimulus distance for positive interpersonal stimuli, achievement-related positive stimuli, negative interpersonal stimuli, and achievement-related negative stimuli (see Dozois & Dobson, 2001b, for more details on scoring). Higher scores indicate greater distance. It was expected that depressive participants would show higher results for positive stimuli, and lower for negative than non-depressive participants. We used four PDST scores to explore if the SSST can detect biased processing in two types of vulnerable participants termed autonomus and sociotopic/dependent (Clark, Beck, & Alford, 1999).

Disfunctional Attitude Scale (DAS; Weissman & Beck, 1978) measures dysfunctional attitudes related to depression (Weissman & Beck, 1978). The scale contains 40 forced-choice items, absolutistic attitudes relating to preconditions for personal happiness. The scale was reported to have satisfactory internal consistency between 0.79 and 0.93 (Olinger, Kuiper, & Shaw, 1987). In line with this, our sample produced 0.81.

---

3 Serbian language has extensive morphological marking and flexible word order, making several solutions possible for each set of words; for this reason the answer key of all possible sentences was reviewed by a professional language editor.

4 The adjectives were translations from the original English version of the test. Following translation, in a pilot study, students for a course credit rated adjectives according to the following criteria: describing social or achievement aspects of the self, valence, frequency of use, and imaginability (Mihic, 2008). Social and achievement adjectives were used given the extant literature on the links between social stimuli, on the one hand, and achievement, on the other, and depression (e.g., Clark, Beck, & Alford, 1999).
Procedure. The research was conducted in 2011 as part of a larger longitudinal study. Group testing was conducted in various faculties of the University of Novi Sad. Participants signed consent forms during their regular classes then were randomly assigned to Group A: load in the first part of the test, n = 563; and Group B: load in the second half, n = 508. The cognitive load – a six-digit number to be remembered – was presented on A0 construction paper, Times New Roman 700, and shown for 30 seconds. All participants had to remember the same number, but a group of senior students proctored them and marked suspicious test protocols. Such protocols were excluded from our study.

Statistical analyses. Each sentence was scored according to two criteria: grammatical correctness and valence. Given that the test was time-limited, uncompleted sentences were treated as incorrect solutions. The following scores were calculated: a) sum of the correct sentences b) sum of the depressive sentences, and c) proportions of the depressive to the correct sentences. The scores were calculated for both load and non-load conditions and were used to examine reliability. In order to explore convergent and divergent validity, the pattern of the correlations between the proportions of depressive sentences, depressive vulnerabilities, and symptom measures was examined.

In order to examine the effectiveness of different testing conditions on elicitation of depressive content, we performed two repeated measures ANOVA-s, on the entire sample and on a sub-sample of individuals scoring either high or low on the DAS scale. In both analyses, the within-subjects factor was presence/absence of cognitive load whereas the between-subjects factor was load order. Depending on our research questions, each analysis included an additional between-subject factor, either vulnerability or gender.5 These analyses were based on proportions of the depressive to the correct sentences, as was the case in the previously reported studies (e.g., Rude et al., 2002; Rude et al., 2003).

Results

Descriptive statistics

Table 1 shows descriptive statistics for the SSST results, expressed as the proportion of depressive sentences to the total number of correctly formed sentences obtained on the entire sample.

<table>
<thead>
<tr>
<th></th>
<th>Depressive sentences with load</th>
<th>Depressive sentences without load</th>
<th>Depressive sentences total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Total</td>
<td>1071</td>
<td>.067</td>
<td>.121</td>
</tr>
<tr>
<td>Males</td>
<td>403</td>
<td>.075</td>
<td>.134</td>
</tr>
<tr>
<td>Females</td>
<td>668</td>
<td>.062</td>
<td>.112</td>
</tr>
<tr>
<td>Group A</td>
<td>563</td>
<td>.064</td>
<td>.123</td>
</tr>
<tr>
<td>Group B</td>
<td>508</td>
<td>.070</td>
<td>.119</td>
</tr>
</tbody>
</table>

Group A – load in the first part of the test; Group B – load in the second part of the test

5 We decided to treat load order as a separate factor even though counterbalancing was used to control for nuisance variables. However, counterbalancing is effective only if there is no interaction between treatment and nuisance variables (Reese, 1997). Given the time-limited nature of the test, there was a possibility that fatigue, practice effects or additional nuisance variables could have interacted with the load order. The only way to explore this was to keep the load order as the between-subjects factor.
Reliability of the SSST

Given that the SSST is a time-limited test, common internal consistency measures could not be calculated. Also, load and non-load conditions could have different effects on test-taking attitudes which additionally complicated reliability analyses. We treated the two blocks of sentences as two parallel forms because halves of the test were administered separately and the timing of each block was reset to zero.\(^6\) The correlations between the two blocks were then corrected by the Spearman-Brown prophecy formula to estimate the reliability of the entire test (Dick & Hagerty, 1971; Fajgelj, 2005). The results were as follows: S-Br between the number of depressive sentences was 0.75; S-Br between the number of correct sentences was 0.79; and S-Br between proportions of the depressive sentences to the total correct was 0.69.

Additionally, Spearman-Brown split-half reliability coefficients (odd and even items) were calculated for items solved with and without load in both halves of the test. Reliability coefficients were lower for the depressive solutions (0.62 with load and 0.64 without load), than for the correct answers (0.93 with load and 0.95 without load).

Gulliksen’s acceleration coefficient for the number of correct answers under cognitive load was 0.31 and without load 0.27. The greater acceleration under load suggests that memorising numbers while solving tasks did indeed place additional demands on cognitive function.

Convergent and divergent validity of the SSST

Table 2 shows the correlations between the production of depressive sentences and symptom measures of depression and anxiety, as well as with measures of vulnerability to depression.

Table 2. Zero-order correlations among the proportions of depressive sentences, symptom and vulnerability measures

<table>
<thead>
<tr>
<th>Instrument</th>
<th>DASS* Anxiety</th>
<th>DASS* Depression</th>
<th>DAS* Disfunc. attitudes</th>
<th>PDST** Achievement (-)</th>
<th>PDST** Achievement (+)</th>
<th>PDST** Inter-pers. (-)</th>
<th>PDST** Inter-pers. (+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depressive sentences – total</td>
<td>.35</td>
<td>.54</td>
<td>.39</td>
<td>-.43</td>
<td>.72</td>
<td>-.48</td>
<td>.72</td>
</tr>
<tr>
<td>Mean</td>
<td>5.57</td>
<td>5.16</td>
<td>11.16</td>
<td>1.28</td>
<td>.34</td>
<td>1.44</td>
<td>.37</td>
</tr>
<tr>
<td>SD</td>
<td>6.03</td>
<td>6.10</td>
<td>22.86</td>
<td>.53</td>
<td>.24</td>
<td>.52</td>
<td>0.25</td>
</tr>
</tbody>
</table>

\(^6\) Although participants worked under different load conditions, these different load conditions were present within each block of sentences, enabling calculation of an approximation of reliability.

*\(^N = 1074; \,**\(^N = 66; \) All correlations were significant, \(p < .001; \) DASS – Depression, anxiety and stress scales; DAS – Disfunctional attitude scale; PDST – Psychological distance scale; Achievement (-) – logarithmic transformation on the PDSTscore for negative achievement attributes; Achievement (+) – logarithmic transformation on the PDSTscore for positive achievement attributes; Interpers. (-) – logarithmic transformation on the PDSTscore for negative interpersonal attributes; Interpers. (+) – logarithmic transformation on the PDSTscore for positive interpersonal attributes.
The proportion of depressive sentences was closely related to the symptoms of depression, although all other correlations were significant, ranging from low to high intensity. The SSST had higher correlations with convergent measures (DASS-depression, DAS and PDST) than divergent measure (DASS-anxiety). The correlation between the SSST and depressive symptoms was significantly higher than between the SSST and anxiety symptoms \((p < .01)\).

In order to test validity of the SSST in relation to the measure tapping cognitive schema structure, we examined correlations between the SSST and PDST, the measure of distances among self-descriptive attributes, on a small sub-sample of participants. Table 2 shows significant correlations between two tests, ranging from medium to high intensity. Production of depressive content on the SSST had higher correlations with the distances between positive attributes than negative, irrespective of their interpersonal and achievement contents.

**Discrimination between vulnerable and non-vulnerable participants**

An additional validity check of the SSST test was conducted on the sub-samples of vulnerable and non-vulnerable participants. We derived a sub-sample of 170 participants whose score on the DAS was greater than \(M + 1SD\), and a group of 156 non-vulnerable participants whose DAS score was \(M - 1SD\) \((M = 110.16 \text{ and } SD = 22.86)\). Descriptive statistics for this sub-samples are given in Table 3.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean (Under load)</th>
<th>SD</th>
<th>Mean (No load)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>326</td>
<td>.095</td>
<td>.157</td>
<td>.089</td>
<td>.143</td>
</tr>
<tr>
<td>Group A</td>
<td>187</td>
<td>.093</td>
<td>.155</td>
<td>.109</td>
<td>.156</td>
</tr>
<tr>
<td>Group B</td>
<td>139</td>
<td>.098</td>
<td>.161</td>
<td>.062</td>
<td>.119</td>
</tr>
<tr>
<td>Non-vulnerable</td>
<td>156</td>
<td>.036</td>
<td>.070</td>
<td>.035</td>
<td>.064</td>
</tr>
<tr>
<td>Vulnerable</td>
<td>170</td>
<td>.149</td>
<td>.192</td>
<td>.138</td>
<td>.174</td>
</tr>
<tr>
<td>Group A*non-vulnerable</td>
<td>83</td>
<td>.036</td>
<td>.078</td>
<td>.048</td>
<td>.077</td>
</tr>
<tr>
<td>Group A*vulnerable</td>
<td>104</td>
<td>.138</td>
<td>.185</td>
<td>.157</td>
<td>.184</td>
</tr>
<tr>
<td>Group B* non-vulnerable</td>
<td>73</td>
<td>.036</td>
<td>.062</td>
<td>.021</td>
<td>.041</td>
</tr>
<tr>
<td>Group B*vulnerable</td>
<td>66</td>
<td>.167</td>
<td>.203</td>
<td>.108</td>
<td>.155</td>
</tr>
</tbody>
</table>

Group A – load in the first part of the test; Group B – load in the second part of the test

The data were subjected to 2x2x2 repeated measures ANOVA: vulnerability and load order (cognitive load in the first or the second half of the test) were treated as between-subjects factors, while presence or absence of cognitive load was treated as a within-subjects factor. Proportions of the depressive sentences to the correct sentences served as the dependent variable. As can be seen in Figure 1, there was a significant effect for vulnerability, \(F(1,322) = 59.64, p < .001, \eta^2_{p} = .16.\) The production of depressive content in vulnerable group was significantly greater than in non-vulnerable group, regardless of load condition.
Vulnerable participants produced 15% of such sentences with load and 14% without, whereas non-vulnerable produced 4% of such sentences both with and without load.

Further analysis suggested a significant interaction between load order and load, \( F(1,322) = 12.71, p < .001, \eta^2 = .04 \). We performed a simple main-effects analysis to better understand this interaction, examining differences in the proportion of sentences with and without load. A Bonferroni correction was applied in order to account for the number of comparisons. Contrasting yielded one significant difference: participants who received load in the second part of the test produced more depressive sentences with load (10%) than without (6%), \( t(138) = 3.45, p < .001 \). Figure 1 shows the main effect of vulnerability and the same interactional pattern between load and load order obtained in vulnerable and non-vulnerable participants. The interaction of all three factors (vulnerability x load order x load) failed to achieve an acceptable level of significance, \( F(1,322) = 2.98, p = .085 \).

**Effectiveness of different testing conditions and gender in eliciting depressive content**

In order to examine effectiveness of different testing conditions data were subjected to 2x2x2 repeated measures ANOVA on the whole sample: gender and load order were treated as between-subjects factors, while presence or absence of cognitive load was treated as a within-subjects factor. The proportion of depressive sentences was a dependent variable. For gender differences, the descriptive results in Table 1 shows that male participants produced a greater proportion of depressive sentences, both with and without load, yielding a
marginally significant gender effect, \( F(1,1067) = 3.57, p = .059, \eta^2_p = .003 \). We checked whether this tendency was connected to depression, and established that the difference between males and females on the DASS depression subscale, \( M_f = 5.31(6.3) \) and \( M_m = 4.90 \) (5.78), was not significant, \( t(1034) = 1.04, p = .30 \).

Analysis also revealed a significant interaction between load order and load, \( F(1,1067) = 20.81, p < .001, \eta^2_p = .02 \), which is shown in Figure 2.

We performed a simple main-effects analysis to better understand the interaction, examining differences in the proportion of sentences with and without load in each group separately. A Bonferroni correction was applied in order to account for the number of comparisons. Differences in both groups were statistically significant: in Group A (load in the first half of the test), the difference was greater for the proportion of sentences without load, \( t(564) = -2.65, p < .05 \); while in Group B (load in the second half of the test), the difference was greater for the sentences with load, \( t(507) = 4.39, p = .000 \). This interaction suggests that both groups produced more depressive content in the second half of the test.

We checked whether the groups with different load orders varied by level of depressive symptoms, which could have affected interpretation of the interaction. However, the difference in depression between groups A and B, as measured by the DASS subscale, was not significant, \( t(1036) = 1.29, p = .20 \).

**Discussion**

*Psychometric characteristics of the SSST*

Using a large sample of students we checked the basic psychometric characteristics of the Serbian Scrambled Sentences Test in various conditions of
administration. The test demonstrated satisfactory internal homogeneity reliability (around .70). Given the nature of the test and the way it was administered, we were able to calculate only a lower-bound approximation of test homogeneity.

Based on the Spearman-Brown split-half reliability, one can see that the test items were less homogeneous in terms of their capacity to elicit depressive responses than grammatically correct responses. Future research should be conducted to explore the following lexical characteristics of the scrambled sentences that could have impacted reliability: frequency of the depressive words and phrases in the vocabulary of the participants, and word order. Some of these lexical features, apart from item content, could have prompted participants to create depressive sentences to a greater/smaller degree. Hence, lexical analyses should guide creation of a shorter version of the test with more homogeneous items, equalized in their capacity to elicit depressive completions.

Construct validity of the SSST was established by examining its correlations with convergent and divergent measures as well as its ability to discriminate between vulnerable and non-vulnerable participants in a theoretically meaningful way. The SSST had higher correlations with convergent measures (DASS-depression, DAS, and PDST) then divergent measure (DASS-anxiety). The correlation of the SSST with depressive symptoms was significantly higher then with symptoms of anxiety, supporting its property as a test of depressive cognitive bias. Also, it discriminated vulnerable from non-vulnerable participants under both load conditions, suggesting that the SSST has discriminative validity as a test of depressive cognitive bias.

The correlation of the SSST with dysfunctional attitudes was significant but low. This result is in accord with other studies (e.g., Rude et al., 2010), in which similar correlations between the DAS and SST were reported (.32 for non-load and .43 for load conditions). Our results, together with Rude et al.’s finding that both DAS and SST contribute to prediction of MDD diagnosis, support the conclusion that these measures tap different aspects of cognitive vulnerability to depression (Rude et al., 2010).

Relation between depressive cognitive bias and structural aspects of depressive cognitive schemas

Psychological distance scaling task is based on the assumption that in depressive and vulnerable persons’ cognitive schemes, compared to non-vulnerable persons, negative self-descriptions are more consolidated, whereas positive self-descriptions tend to be more disperse. Because negative self-descriptions are more interconnected, activation of one such attribute activates related attributes more effectively (Dozois & Dobson, 2001b). Our results support the argument that structural and process aspects of depressive cognitive organization are closely related.

---

7 Within each block of sentences, there was an equal number of sentences in which positive and negative words appeared first. Also, within each block, the order of sentences was randomized. However, it is possible that some wordings were more in tune with the way our cognitive system processes language making some depressive completions more difficult/easier than others.
The greater diffusion of positive attributes in the scheme, rather than connectedness of negative attributes, appeared to be highly related to negatively biased solutions to the scrambled sentence test. From this, it may be concluded that the production of depressive sentences is largely enabled by the lower availability of positive content in the memory, rather than by the higher availability of negative content. This conclusion, however, should be considered cautiously because of the correlational nature of the study, the sample from general population, and incongruity between SSST and PDST items. To our knowledge, this is the first study in which these two measures were compared, highlighting the need for replication studies. The only data that can be related to our finding is that of Dozois and Dobson (2001a) who found that symptoms of depression on the Beck Depression Inventory were also more related to positive PDSQ distances then negative. They also found that positive cognitive organization breaks down earlier during development of dysphoria and tends to recover earlier then negative during remission.

Cognitive bias and suppression in vulnerable and non-vulnerable participants

According to Ironic Process Theory, vulnerable participants, compared to non-vulnerable, would demonstrate a greater proportion of depressive sentences only under cognitive load. However, vulnerable participants in our study demonstrated a greater depressive bias under both testing conditions compared to non-vulnerable. A possible explanation can be found in a non-transparent and cognitively demanding nature of the whole test. Similarly, Rude et al. (2002) have argued that the test, even without load, elicits depressive content in participants because it is presented as a skill test with a time limit which can shift focus away from depressive thoughts and weaken distraction in participants.

Another departure from Ironic Process Theory and the previous studies was that even within the non-load condition, vulnerable participants reported more negative sentences compared to non-vulnerable. A reason can be found in the definition of at risk individuals. Previous studies (e.g., Wenzlaff & Bates, 1998) defined vulnerable participants as those with a history of depression whereas in our study the vulnerability status was determined based on self-reports on the DAS scale.

The result which is partly in accord with the theory (i.e., greater proportions of depressive content with load compared to non-load) was obtained only in participants who were given memorising task in the second half of the test. It is possible that the combined effects of fatigue and cognitive load led to more depressive sentences. This result was obtained on the entire sample and also on the vulnerable and non-vulnerable sub-samples, suggesting that suppressive tendencies might be present in our non-vulnerable participants too. One explanation for such findings may lie in the self-descriptive nature of the vulnerability measure (i.e., the DAS), which is susceptible to deliberate attempts to present the self in a socially favourable light.
Effect of load order and load on cognitive bias and suppression

The three testing conditions (load in the first half of the test, load in the second half, and non-load in the second half) yielded approximately equal and higher proportions of depressive sentences compared to the condition of non-load in the first half. This result could not be attributed to potential differences in current dysphoric symptoms among groups with different load order. The unexpected result that the highest proportion of depressive sentences was produced under non-load condition in the second part of the test, may indicate rebound effect (Najmi & Wegner, 2009). In the group which already had load in the first half of the test, early fatigue may have weakened mental control, and ability to suppress in the second half. The increased effort of maintaining suppression in the first half may facilitate the infiltration of depressive content during the second half.

Gender differences in the production of depressive content on the SSST

Our results suggested a trend toward gender differences in the production of depressive sentences. Male participants tended to produce more depressive sentences than females, both with and without load. This result was interesting because males showed no significant difference from females in levels of depression, and indeed showed slightly lower averages for depression than females. Although gender differences were only marginally significant, this trend should not be discounted completely given that it is in accordance with the previous studies (e.g., Rude et al., 2002).

Limits of the research

Although we were able to derive a sufficiently large group of vulnerable participants, a question remains about the sufficiency of the DAS scores as the sole criterion for establishing vulnerability. Better criterion for detecting vulnerability would be a history of depression either personally or in the immediate family. Predictive validity should be tested in prospective designs, in which is possible to find out how vulnerability measured by the SSST predicts depressive symptoms in non-clinical and clinical samples.

Furthermore, although the DAS is widely used to measure vulnerability to depression, hypotheses stemming from Ironic Process Theory would be tested better if suppression tendencies were directly measured (e.g. by means of the White Bear Suppression Inventory; Wegner & Zanakos, 1994).

The greatest shortcoming of this research is the lack of counterbalancing of blocks of sentences. Although the sentences were randomized within blocks, and the load order was counterbalanced, the same group of sentences appeared only in the first or second half of the test. As a consequence, one possible explanation for the greater production of depressive content in the second half of the test may lie in the wording of the stimuli themselves. To our knowledge, this lack of counterbalancing of the sentences is common across various foreign studies with the SST. Researchers counterbalanced only load order (eg. Rude, et al., 2010; Watkins & Moulds, 2007), assigned randomly participants to load or non-load condition (e.g.Van der Does, 2005), or assigned all participants to
a single, load condition (e.g. Holmes et al., 2009). However, our results suggest that counterbalancing of sentences is also necessary to better understand cognitive processes provoked by SST.

**Conclusions**

The SSST was shown to possess satisfactory psychometric properties and to be essentially similar to the instrument used by American authors. Multiple analyses (relations to depressive symptoms and measures of vulnerability tapping dysfunctional attitudes and structural aspects of cognitive schemas) showed that the SSST is a measure of depressive bias. As a measure of suppression, the SSST performed as expected only partly when load was applied in the second half of the test. The construct validity of the test was confirmed, not only via the specific relation with depressive symptoms and the ability to distinguish vulnerable and non-vulnerable participants, but also via the significant and stronger relation to structural aspects of the depressive cognitive schema. However, some specific hypotheses stemming from Ironic Process Theory were not supported, suggesting a need for further research which would examine the effects of various nuisance effects such as fatigue, word order and frequency.

The advantages of the SSST, demonstrated by our research, over other tests of interpretive bias are as follows:

- SSST scores behave in accordance with Beck’s cognitive schemas theory, supporting construct validity of the test;
- Ease and speed of administration, making the instrument suitable for both research and clinical practice;
- Ability to discriminate depressive and vulnerable persons while the true nature of the test is not disclosed to the examinees.

**References**


