SOCIAL SCIENCES & HUMANITIES

Trust your Abilities More than the Stereotype: Effect of Gender-stereotype Threat and Task Difficulty on Word Production, Clustering, and Switching in Letter Fluency

Sri Wulandari and Donny Hendrawan*

Faculty of Psychology, Universitas Indonesia, Depok 16424, Indonesia

ABSTRACT

Gender-stereotype threat consistently accounts for underperformance phenomena experienced by women on male-stereotyped cognitive tasks. However, only a few studies have examined how the threat is affecting performance on female-stereotyped cognitive tasks, such as letter fluency. The present study examined whether variations in the cues to activate stereotype threat and the level of task difficulty would affect the letter fluency performance of undergraduate men and women ($n = 168$) and the underlying cognitive processes of this performance (i.e., switching, clustering). The results indicated participants held beliefs about women's superiority in this task. However, threat-activation cues did not affect production of correct words, errors, clustering, or switching in men and women. Task difficulty affected the number of correct words, yet it did not interact with the stereotype threat-activation cues. Finally, participants' actual performance was related to their self-rating perception about their ability instead of the stereotyping they perceived. The effect of self-efficacy, educational level, and individuals' susceptibilities should be taken into account when studying the effects of stereotype threat.

Keywords: Clustering, gender-stereotype threat, letter fluency, task difficulty, switching

*E-mail address*es:
sri.wulandari71@ui.ac.id (Sri Wulandari)
donny.hendrawan@ui.ac.id (Donny Hendrawan)
*Corresponding author

INTRODUCTION

Gender-stereotype threat refers to the phenomenon in which members of a particular gender group fear that their performance on a particular task will confirm an existing negative stereotype attached to their gender and this leads to a poorer performance in that stereotyped task (Spencer et al., 1999; Steele, 1997). Studies of gender-stereotype threat have mostly examined the effects on gender-sensitive tasks favouring men, such as mathematics (Flore et al., 2018) and spatial ability (Sanchis-Segura et al., 2018). By introducing a gender-stereotype threat into a testing situation, score differences between gender groups are produced or enlarged, with men consistently outperforming women. Despite the growing number of studies on gender-stereotype threat (Spencer et al., 2016), only a few of them examined the effect of a gender-stereotype threat on gender-sensitive tasks favouring women, such as social sensitivity (Koenig & Eagly, 2005), reading performance (i.e., recognition and comprehension of written words; Pansu et al., 2016), or verbal fluency (Hirnstein et al., 2014; Hirnstein et al., 2012).

A study using a social sensitivity task showed that the male participants in the threat condition performed worse than both the female and the male participants in the non-threat condition (Koenig & Eagly, 2005). In this study, gender stereotype threat was activated by explicitly telling the participants that men usually performed worse on the task and that the task was meant

to measure people's ability to understand communication and the use of nonverbal cues. Additionally, a study utilizing reading performance tasks (Pansu et al., 2016) employed a subtler method of activating gender stereotype threat by telling students in the experimental group that the task was a test meant to evaluate their reading ability. As expected, the male students performed worse in the threat conditions than the female students. In contrast, despite the stereotype about women's superiority in language-related domain (Moe et al., 2020; Steffens & Jelenec, 2011), studies using verbal fluency found that gender-stereotype threat did not necessarily cause the performance of men in the threat group to be worse than men in the control group (Hausmann et al., 2009; Hirnstein et al., 2012; Hirnstein et al., 2014).

Although studies regarding verbal fluency found that men did not perform worse when under stereotype threat (Hausmann et al., 2009; Hirnstein et al., 2012; Hirnstein et al., 2014), it was found that the extent of the threat effects on the participants' performance varied among these three studies based on the way that the threat was activated. Hirnstein et al. (2012) found that when the threat was activated in a moderately explicit manner (i.e., participants were told that they would participate in a study about gender differences using verbal tests), the men in the threat group performed better than those in the control group. Additionally, Hausmann et al. (2009) and Hirnstein et al. (2014) found that the

use of subtle cues to activate the stereotype threat (i.e., participants were asked to complete a gender-stereotype questionnaire to prime the stereotype threat before the letter fluency test was administered) did not cause the performance of the men in the stereotype-threat conditions to be significantly different from that of the men in the control group. Meta-analyses also supported that the way in which the stereotype threat was activated moderated the effect of gender stereotype threat on the participants' performance (Appel et al., 2015; Nguyen & Ryan, 2008; Walton & Cohen, 2003). However, the results from these meta-analyses still yielded different conclusions regarding which activation cue would produce the largest effect size. Hence, a study in which the three activation cues are directly compared is needed.

Another important variable that significantly affects participants' performance is the level of task difficulty (Flore et al., 2018; Spencer et al., 2016). According to Spencer et al. (2016), gender-stereotype threat only interferes with participants' performance when the task is considered as difficult. For example, in the spatial orientation task, a significant effect of gender-stereotype threat was present in males and females in a more difficult task, but not in an easy one (Allison et al., 2017). When the difficulty level was reduced, both male and female showed better performance even under the stereotype threat. The findings on the level of difficulty are also supported by a meta-analysis that showed a robust moderating effect of task difficulty; the more difficult the task, the larger the effect size produced by the stereotype threat (Nguyen & Ryan, 2008).

According to Spencer et al. (2016), introducing a gender-stereotype threat during an easy task can improve individuals' performance due to their heightened motivation to disconfirm or avoid confirming the stereotype. However, when a task is challenging and demanding, the cognitive demand of the task increases individuals' mental workload, making interference from a stereotype threat cognitively more problematic (Steele et al., 2002). Stereotype threat interferes with test-takers' cognitive performance by using their working memory capacity to suppress negative thoughts about the stereotype threat, i.e., the thought that their performance will confirm an existing negative stereotype about their social group (Beilock et al., 2007). As a result, working memory capacity that is available to attend to a task's information-processing requirements and controlling its execution is reduced (Schmader et al., 2008), and the performance becomes compromised.

While it has been evident that under stereotype-threat condition, level of task difficulty affects participants' performance on mathematics and spatial task, to date, no study on stereotype threat has incorporated task difficulty to examine how it affects performance on cognitive tasks favouring women. Letter fluency is one of the women's stereotypical tasks that relies on working memory (Azuma, 2004; Mueller et al., 2015)

and is also relatively easy to be manipulated in terms of difficulty level. In this task, participants should generate words begin with specific letters. The level of difficulty depends largely on the specific letter used as the stimuli; as the difficulty increases, the number of correct words decreases, and the errors increase (Borkowski et al., 1967; Kosmidis et al., 2004).

Regarding the effect of stereotype threat on performance, there are a number of psychological factors that moderate the degree of individuals' susceptibility to a gender-stereotype threat. These moderators affect the extent to which individuals are invested in the evaluative aspect of the stereotype and the extent to which they are able to cope with the pressure from the stereotype (Spencer et al., 2016). According to Nguyen and Ryan (2008) and Spencer et al. (2016), individuals tend to invest more in the success of their performance when they have a high level of stigma consciousness (i.e., the extent to which individuals are conscious of the negative stereotypes attributed to their social group), domain identification (i.e., how strongly individuals identify with the domain being evaluated), and group identification (i.e., how strongly individuals identify with the social group being stigmatized).

Another psychological factor, namely, self-efficacy belief, could also affect individuals' ability to preserve their performance under the additional pressure applied by the stereotype. Self-efficacy belief is defined as a person's belief about his/her ability to succeed in a particular task despite the circumstances (Bandura, 1997). It has been argued that self-efficacy is critical due to its positive influences on performance as well as its role in mitigating the negative influences on performance (Bajaj et al., 2014); it does this by focusing individuals' attention, monitoring their execution of tasks (Mayer & Moreno, 2003), and enabling them to resist the distractions caused by anxiety and handicapping (Ashcraft & Kirk, 2001; Hopko et al., 2005). Moreover, individuals may approach a task differently based on their self-efficacy level and may adopt different cognitive strategies with different levels of efficiency (Kalpouzos & Eriksson, 2013). Individuals with higher levels of self-efficacy are found to be more able than others to sustain their performance when a task becomes complex and requires more working memory (Hoffman & Schraw, 2009). Individuals' levels of self-efficacy and those of their susceptibility should be taken jointly into consideration when studying the effect of stereotype threat, as these factors may serve as a buffer or a catalyst for individuals' performance.

The letter fluency performance of the adult sample was found to be affected by several factors, such as educational level (Aziz et al., 2017; Ghasemian-Shirvan et al., 2018), level of literacy, ethnicity (Strauss et al., 2006), and intelligence (Diaz-Asper et al., 2004). Meanwhile, the findings regarding gender and age were quite mixed (Ghasemian-Shirvan et al., 2018; Mathuranath et al., 2003; van Der Elst et al., 2006; Zimmerman et al., 2014). The studies performed in Indonesia that

used undergraduate students as a sample showed that university affiliation (i.e., state university vs private university), intelligence, and creativity fluency affected letter fluency performance, while gender and academic background (i.e., exact science vs social science) had no effect (Hendrawan, 2013; Hendrawan et al., 2015). Hence, it is important to control these demographic and ability-related variables when a letter fluency task is used as a measure of cognitive performance, such as by ensuring that the sample is homogeneous in terms of participants' levels of education, literacy, intelligence, or age groups.

In letter fluency test, optimal overall performance is characterized by a high number of correct words and a low number of errors (Strauss et al., 2006). However, in neuropsychological task such as letter fluency, it is equally important to study both the overall performance and the underlying strategies used by individuals to reach the optimal performance (Abwender et al., 2001). This optimal performance in letter fluency is affected by two underlying cognitive processes, i.e., clustering and switching (Troyer et al., 1997). Clustering is defined as the ability to generate phonemically related words within a subcategory, while switching is defined as the ability to shift to a new subcategory when the previous subcategory is exhausted (Troyer, 2000).

Studies on clustering and switching have shown that men and women rely on different strategies to generate words; men tend to produce significantly more clusters than women (Lanting et al., 2009; Weiss et al., 2006), but women tend to switch more than men (Weiss et al., 2006). However, both strategies appeared to be successful in producing an optimal performance because both men and women created a comparable number of total correct words (Lanting et al., 2009; Scheuringer et al., 2017). Given that an optimal performance in letter fluency is affected by two different word generation strategies, and different sexes employed different strategies, we extend our present study by examining whether the effect of a gender-stereotype threat is evident in the component processes of letter fluency performance, i.e., clustering and switching.

Taken together, considering that (1) less consistent results are evident on gender stereotype-threat studies using the letter fluency task, (2) stereotype threat-activation cues as well as level of task difficulty may influence the effect of stereotype threat on individuals' performance, and (3) the effect of a gender-stereotype threat may be observed not only in the overall performance but also in the component processes of letter fluency test, this study aimed to address how different gender stereotype threat-activation cues and different levels of task difficulty would affect men and women on letter fluency task performances in the aspects of number of correct words, errors, clustering, and switching. We hypothesized that different types of gender-stereotype activation cues would affect male and female performance differently, and increasing difficulty levels will decrease their letter fluency performance.

MATERIALS AND METHODS

Participants

A total of 168 undergraduate students (91 female) from a state university in Depok, Indonesia, were recruited. Participants were excluded from this study if they did not speak Bahasa as their first language for daily communication, had a history of trauma and/or surgical operation related to the brain and central nervous system, suffered from neurological and psychiatric disorders, had a history of substance and/or alcohol abuse, or were left-handed or ambidextrous. Given that letter fluency is a neuropsychological test, these criteria were applied to ensure that language barrier and neurological or psychiatric conditions did not affect participants' letter fluency performance.

Study Design

This study was a single-blind, randomized and controlled experiment. Participants were randomly assigned to four gender stereotype-activation groups (blatant/moderately explicit/subtle/no activation with the number of participants being 44, 41, 42, and 41, respectively) and assessed three times as the level of difficulty in the letter fluency test increased (easy, medium, and hard).

Materials

Gender Stereotype-activation Cues. Gender stereotype threat-activation cues were pre-recorded using Audacity® 2.0.3 for Windows. The instructions and procedure for activating gender-stereotype threat were adapted from Nguyen and Ryan (2008), and Hirnstein et al. (2012). In the control group, participants were told that they would engage in a task that was a part of a research about the general process of problem solving. In the subtle group, participants were told that they would take a verbal fluency test to evaluate their language ability. In the moderately explicit group, participants were told that their language ability would be evaluated and that gender difference existed in verbal fluency performance so that males and females would get a different score. In the blatant group, participants were told that they would be evaluated on their language ability and women would get a higher score than men in this verbal fluency test. Apart from the control group, participants were asked to indicate their gender and to write it down on a piece of paper.

Letter Fluency Test. Letter fluency performance was measured by a letter fluency test that consisted of three letters, i.e., K, R, W; each letter represents easy, moderate, and hard level of difficulty (Hendrawan & Hatta, 2010; Hendrawan et al., 2015). The instructions and test items were pre-recorded. Four scores were obtained for each participant: (a) number of correct words, (b) number of errors, (c) mean cluster size, (d) number of switches. Scoring was completed by giving +1 for every correct word and 0 for every error. The score for the correct word/ error was calculated by summing up the total number

of correct words/ errors produced for each letter. A word was considered an error if it started with a capital letter (e.g., name of persons, places, or brands), was a repeat, was meaningless, or differed only by its suffix. The online Indonesian dictionary (http://kbbi.web.id/) was used to check whether the words were officially recognized in Bahasa. A cluster was defined as a group of words sharing similar characteristics, i.e., begin with the same two first letters, rhymed, differed only by a vowel sound, or were homonyms. Cluster size was counted from the second word in each cluster. The mean cluster size was computed by dividing the total cluster size by the total number of clusters for the three letters. The number of switches was defined as the number of transitions between clusters. Error words were included when performing the scoring for clustering and switching. Detailed rules for scoring clustering and switching can be found in Troyer et al. (1997).

Gender-stereotype Questionnaire. The gender-stereotype questionnaire was used to determine the direction and magnitude of gender stereotype held by the participant regarding verbal ability. The questionnaire had been translated into Bahasa from the gender-stereotype questionnaire by Hausmann et al. (2009) and permission was obtained from the authors. We only used one item from this questionnaire that refers to verbal fluency, as this was the focus of the present study. Participants were asked to imagine that they would meet someone whom they had never met before, and they received a clue that this person "…*can generate many words beginning with the same letter within one minute"*. Based on that clue, participants rate the probability that this person was man or woman. To indicate their rating, participants were asked to put any number in the male and female columns as long as the sum was 100. A probability estimate of 50% in the male column and 50% in the female column meant that participants believed there was no gender difference in the word generation ability mentioned in the questionnaire. An estimate larger than 50% in the male column indicated that the participant believed the person in the story was more likely to be male and vice versa.

Self-rating. The self-rating was also translated from Hausmann et al. (2009). The self-rating was given to the participants to assess the possibility that although participants might believe that letter fluency ability was generally associated with one gender or another, as an individual, their ability on letter fluency was an exception and was independent from the stereotype. Participants were asked about how they perceived their ability on the letter fluency test they had just done. To complete the self-rating, the participants were asked, "*To what extent does the following item describe you? I can generate many words beginning with the same letter within one minute"*. On a 7-point scale, the participant was asked how applicable that statement was to them, with '1' indicating 'not descriptive at all' and '7' indicating 'highly descriptive'.

Manipulation Check

A manipulation check was completed by giving a questionnaire asking whether the recorded instruction for activating a gender-stereotype threat was perceived as having a different degree of stereotype across the three manipulation groups (blatant, moderately explicit, subtle) and a control group. The instruction on how to fill this manipulation check was similar to the gender-stereotype questionnaire.

Procedures

Informed consent was obtained from all participants. Each participant was randomly assigned into one of four gender stereotype threat-activation groups. An experimenter played one of the pre-recorded stereotype-activation cues using a laptop (Windows media player on an HP Pavilion dm1) followed by the pre-recorded letter fluency test. The test consisted of three parts, i.e., test instructions, sample item, and test items. The experimenter took notes every word generated by the participants. After the test finished, the participants were given gender-stereotype questionnaire and self-rating. The experimental procedure was carried out in accordance with the Declaration of Helsinki.

Statistical Analysis

Scores from manipulation check violated the assumption of normality and heterogeneity; hence Kruskal-Wallis Test was used to examine how gender stereotype threat-activation cues were perceived across the four activation groups whilst Kolmogorov-Smirnov Test was used to examine across gender. To examine the effect of activation types, gender, and level of task difficulty on letter fluency performance, number of correct words and errors from each letter on the letter fluency test were subjected to 4 x 3 x 2 mixed ANOVA. The total correct words, total errors, mean cluster size, and number of switches from the letter fluency test were subjected to 4 x 2 ANOVA to examine the effect of activation types and gender on letter fluency overall performance and its component processes.

One sample t-test was performed on the probability estimate from the male column of the gender-stereotype questionnaire to examine whether the probability estimation was significantly different from a value of 50 (an indication of neutral appraisal). A significantly positive t-value indicated the probability as a male. A one sample t-test was also performed on the self-rating score to examine whether the score was significantly different from a value of 4 (an indication neutral appraisal of their performance). A significantly positive t-value indicated a generally superior appraisal. To further examine whether gender stereotype and performance appraisal varied across activation type and gender, probability estimates from the gender-stereotype questionnaire and scores from self-rating were subjected to 4 x 2 ANOVA.

Pearson correlations were performed on the scores from the letter fluency test (number of correct words, number of errors, mean cluster size, number of switches), gender-stereotype questionnaire, and self-

rating to examine whether the scores from these instruments correlated with each other.

RESULTS

Manipulation Check

The median score from the male column on the manipulation check showed a gradual increase across the four stereotype threat-activation groups, with the lowest mean produced by the blatant group, followed by the moderately explicit, subtle, and control group, as shown in Figure 1. The median score from the male column on manipulation check showed a comparable value between the 2 gender groups. The instruction was significantly different across activation groups, $H(3) = 25.959$, $p = 0.00$, but not significantly different across gender. Pairwise comparisons with adjusted p-values showed that there were significant differences in the score from the male column between the blatant activation group compared to the rest, as shown in Table 1.

Letter Fluency Test. The participants' scores on the letter fluency test are presented in Table 2. As shown in Figure 2, analysis of correct words showed that the level of task difficulty was the only variable with a significant effect on the letter fluency performance, as the difficulty increased the number of correct words decreased, $F(2, 320) = 760.14$, $p < 0.05$, $\eta^2_P = 0.826$. Neither gender stereotype-threat activation nor gender showed a significant effect and no interaction effect existed between the three variables. When number of error words for each letter stimulus was entered as a dependent variable, analysis showed that gender-stereotype threat, gender, or task difficulty had no significant effect.
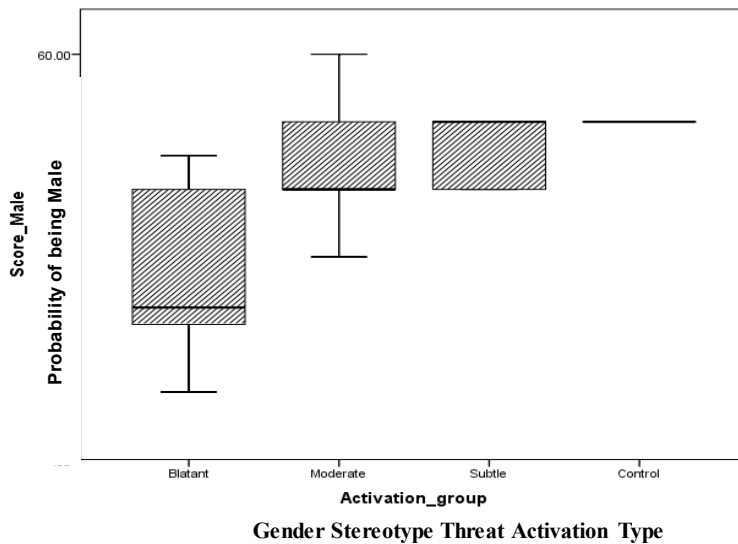

Figure 1. Differences on scores from male column across four gender stereotype activation groups

Table 1

Pairwise comparisons between blatant, moderately explicit, subtle, and control activation group

Activation Type		1	2	3	4
1	Blatant	-	$H = -16.57$ $p = 0.017*$, $r = 0.433$	$H = -20.25$, $p = 0.002*$, $r = 0.544$	$H = -28.68$, $p = 0.000*$, $r = 0.80$
2	Moderately Explicit	-	-	$H = -3.68$, $p = 1$, $r = 0.151$	$H = -12.11$, $p = 0.245$, $r = 0.138$
3	Subtle	-	-	-	$H = -8.43$, $p = 0.927$, $r = 0$
4	Control	-	-	-	-

Notes: Test Statistic (H), Adjusted P-value (p), Effect Size (r), *Significant at p < 0.05

Table 2

Mean and standard deviation between sex groups and activation groups on letter fluency scores

Letter Fluency *M*(SD)	Sex		Activation Groups			
	Male	Female	Blatant	Moderate	Subtle	Control
K/ K-Error	17.71 (3.91)/ 0.68 (0.83)	18.58 (3.97)/ 0.74 (1.02)	18.20 (4.65)/ 0.82 (1.06)	18.20 (3.83)/ 0.85 (0.96)	17.62 (3.64)/ 0.57 (0.77)	18.73 (3.64)/ 0.59 (0.92)
R/ R-Error	13.21 (2.86)/ 0.58 (0.98)	13.90 (4.33)/ 0.68 (0.87)	13.77 (3.88)/ 0.43 (0.66)	12.51 (3.43)/ 0.78 (1.29)	13.64 (3.90)/ 0.69 (0.81)	14.39 (3.56)/ 0.66 (0.79)
W/ W-Error	6.74 (2.62)/ 0.58 (0.89)	7.10 (2.38)/ 0.60 (0.92)	6.89 (1.98)/ 0.59 (0.73)	6.56 (2.13)/ 0.83 (1.16)	6.83 (3.37)/ 0.40 (0.80)	7.46 (2.26)/ 0.56 (0.87)

Table 2 *(Continued)*

Letter Fluency M(SD)	Sex		Activation Groups			
	Male	Female	Blatant	Moderate	Subtle	Control
Total words/ Total error	37.66 (7.71)/ 1.83 (1.88)	39.60 (8.40)/ 2.02 (1.75)	38.86 (8.93)/ 1.86 (1.62)	37.27 (7.53)/ 2.41 (2.30)	38.14 (8.77)/ 1.67 (1.51)	40.59 (6.97)/ 1.80 (1.68)
Mean cluster size	0.85 (0.59)	0.70 (0.37)	0.78 (0.42)	0.69 (0.43)	0.78 (0.49)	0.82 (0.60)
Switching	20.07 (7.34)	22.07 (6.46)	21.23 (8.01)	21.49 (6.45)	20.67 (6.99)	21.25 (6.23)

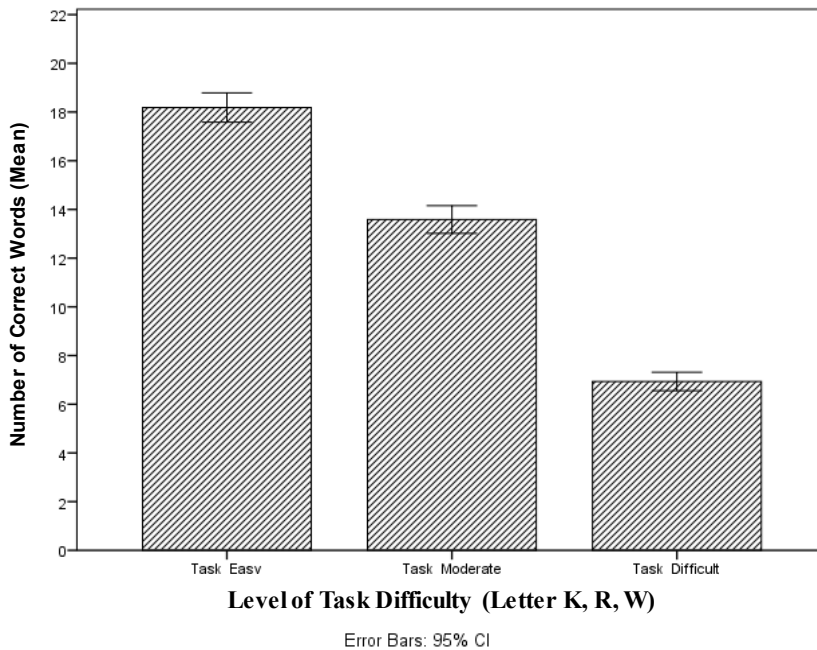Notes: Mean (M), Standard Deviation (SD)

Figure 2. Differences on number of correct words across three levels of task difficulty

Analysis of mean cluster size showed that gender had a significant effect on mean cluster size $F(1,159) = 4.123$, $p < 0.05$, $\eta^2_P = 0.025$, with male participants having a higher mean value than female participants ($M = 0.849$, $M = 0.697$, respectively). Gender stereotype-threat activation did not have a significant effect on mean cluster size. Analysis of the number of switches showed that neither gender nor gender stereotype-threat activation had a significant effect on the number of switches.

Gender-stereotype Questionnaire and Self-rating. The mean probability estimate from the male column across all participants was $M = 46.16$ ($SD = 12.74$) and differed significantly from a value of 50, $t(167) = - 3.91$, $p < 0.05$, $d = -0.30$, indicating that in general, participants held a gender stereotype favouring women. The mean probability estimates from male and female participants were also calculated separately. The mean probability estimate from men was $M = 41.56$ ($SD = 11.40$) and differed significantly from the value of 50, $t(76) = -6,50$, $p < 0.005$, $d = -0.74$, indicating that men believed that the hypothetical person with letter fluency ability was more likely to be a woman. In other words, they believed that letter fluency task was more likely to have a gender-stereotype favouring women. The mean probability estimate from women was $M = 50.05$ ($SD = 12.57$) and did not differ significantly from the value of 50, $t(90) = 0.042$, $p > 0.05$, $d = 0.00$,

indicating that women did not believe a gender stereotype favouring women existed in letter fluency task. In other words, women believed that letter fluency task was more likely to be gender-neutral. The 4 x 2 ANOVA for the probability estimate showed that there was a significant effect of gender, $F(1,160) = 21.62$, $p < 0.01$, $\eta^2_P = 0.119$, but there was no significant effect of stereotype threat-activation type.

A similar procedure was performed on the self-rating score. Even though in general participants rated themselves lower than the average score ($M = 3.87$, $SD = 1.09$), one-sample t-test showed that the overall self-rating score and gender-based self-rating scores (men and women scores calculated separately) did not differ significantly from a value of 4, which was the average score. The 4 x 2 ANOVA for the self-rating score showed that gender stereotype activation type and gender had no significant effects.

Correlations between Total Correct Words, Total Errors, Mean Cluster Size, Number of Switches, Gender-stereotype Score, and Self-rating Score. As shown in Table 3, significant positive correlations were obtained between total correct words and number of switches, total correct words and the self-rating score, total errors and the gender-stereotype score, total errors and the self-rating, mean cluster size and number of switches, and number of switches and the self-rating score.

Table 3

Correlations between total correct words, total errors, mean cluster size, number of switching, gender stereotype score, and self-rating score

Score	1	2	3	4	5	6
Total Correct Words	-	.045	.066	.590**	-.100	.223*
Total Errors	-	-	.007	.135	.222**	.160*
Mean Cluster Size	-	-	-	-.666**	-.119	-.081
Number of Switching	-	-	-	-	.017	.238**
Gender-stereotype	-	-	-	-	-	.052
Self-rating	-	-	-	-	-	-

Notes: *Significant at $p < .05$, ** Significant at $p < .01$

DISCUSSION

This study aimed to examine whether different gender stereotype threat-activation cues would affect men's and women's performances differently when tested over three levels of difficulty on the letter fluency test. Overall, this study found that, although the results of the gender stereotype questionnaire showed that the participants held certain beliefs about the presence of a gender stereotype in the letter fluency test, participants' actual performance was not affected by the stereotype threat. This result was still the case even when we accounted for the variation in the stereotype threat activation cues and the level of task difficulty. Regarding the variations in the threat activation, our analysis of the manipulation check showed that participants perceived that the cues that activated the gender stereotype threat were gradually increased in terms of emphasizing women's superiority. In terms of task difficulty, the analysis showed that the level of task difficulty affected the production of the correct words but did not affect the errors for each letter stimulus. However, even after the cognitive demand of the test was manipulated by introducing three different letter stimuli with an increased level of difficulty, the participants' performance under the stereotype threat was not affected.

Interestingly, it was found that the participants' performance in this test was related to their self-rating, which measured their perception about how well they performed on the letter fluency test. The scores from the participants' self-ratings accurately reflected the participants' actual performance on the letter fluency test. Furthermore, when the variations in gender and stereotype-threat activation were

considered, the scores from the participants' self-ratings showed that these scores were not affected by the variations in gender and stereotype-threat activation cues. This result regarding the self-rating scores was in line with the participants' scores on the letter fluency test, which were also not affected by the variation in gender or stereotype-threat activation cues. Collectively, our results showed that neither the participants' perceptions about women's superiority in relation to the activation cues nor their beliefs about the gender stereotypes attached to the task necessarily caused differing performances among the participants during the letter fluency test. The participants' own judgement about their ability was more important for their performance than the abovementioned factors (Hausmann, 2014).

This study also examined the effect of the gender-stereotype threat on clustering and switching as the underlying cognitive processes of the letter fluency test. We found that clustering and switching were not affected by the gender-stereotype threat. When we analysed clustering and switching between gender groups, we found that men produced significantly larger cluster sizes than women. In terms of switching, the two genders produced a comparable number, although men switched slightly less than women. Our findings were also partially consistent with Troyer et al. (1997) and Scheuringer and Pletzer (2017); we found that mean cluster size and the number of switches were negatively correlated, and switching was more highly related with total correct words as well as with the self-rating

score. Our findings suggested that men and women relied on different strategies to generate words, but both strategies worked similarly well in producing total correct words. Although males tended to produce more words within a cluster, the mean cluster size was not related to the total correct words as much as switching was. Hence, when females compensated for smaller cluster sizes by being more rapid in switching, the total number of correct words between two genders was relatively comparable.

Based on our findings, we propose that self-efficacy, educational level, and individual moderators may explain on why gender-stereotype threat did not affect letter fluency performance even when the task difficulty had been increased. It is important to highlight the role of participants' self-efficacy belief (Bandura, 1997), given that our data showed that self-rating reflected participants' actual performance on letter fluency better than the belief about the stereotype threat. A study by Chung et al. (2010) showed that self-efficacy mediated the effect of ethnicity-stereotype threat on a job knowledge exam. This result is understandable, as self-efficacy has been found to exert positive influences and mitigate the negative impacts on performance in a pressured situation, such as when an individual is under stereotype threat (Bajaj et al., 2014; Pennington et al., 2016). Previous studies also found that participants with high self-efficacy employed more efficient cognitive strategies than participants with low self-efficacy;

hence, these participants performed better on cognitive tasks, such as set-shifting (Mäntylä et al., 2010) and visuospatial working memory tasks (Kalpouzos & Eriksson, 2013).

In our study, scores from the self-rating showed that participants' beliefs about their performances on the fluency test were robust. Across different stereotype-threat conditions, participants held similar beliefs that their letter fluency ability was average. We argued that this positive belief enabled participants to block the negative influence and simultaneously enabled them to employ more efficient cognitive strategies to sustain their performance even when the test became more demanding for their working memory due to the increased level of difficulty.

Another important aspect is the participants' educational level as a university student. According to Andreoletti and Lachman (2004), level of education buffers the effect of the gender-stereotype threat on cognitive performance. This study found that under stereotype-threat conditions, individuals with less than a 4-year college degree had a lower score on a memory test, while individuals with higher education performed better on the same test. It was suggested that higher level of education is associated with better use of adaptive memory strategies and with greater confidence and self-control (Andreoletti & Lachman, 2004; Schieman, 2001). Furthermore, it is also worth mentioning that the research on verbal fluency among Indonesian undergraduate students also found that university affiliation affected participants' verbal fluency performance (Hendrawan et al., 2015); the state university students produced a higher number of correct words in the letter fluency tests than the private university students. In Indonesia, it is well known that state universities apply more rigorous admission criteria and have a more competitive admission rate; hence, this group difference could be due to the relatively different cognitive abilities between these two groups (Hendrawan et al., 2015).

In our study, participants' performances in the letter fluency test showed that their production of correct words and errors were not affected by the gender-stereotype threat even after we varied the level of task difficulty. Moreover, participants produced a very low number of errors, leading to a very positively skewed distribution with a narrow range of scores, suggesting a floor effect. Taken together, it is reasonable to argue that, for our participants, performing a letter fluency test under the presence of stereotype threat was not cognitively problematic given their educational background (i.e., university students who came from one of the best state universities). Their advantageous educational background means that they possess above average cognitive capacities that enabled them to employ better coping strategies. Furthermore, better educational background might also contribute toward participants' level of self-efficacy which then helped them to sustain their performance under the stereotype threat.

Underlining the highly situational

character of stereotype threat (Steele, 1997), it has been shown that the degree and direction of stereotype threat effects can be moderated by various psychological factors (Appel et al., 2015; Spencer et al., 2016). Among these psychological factors, the three most important individual moderators which cause susceptibility towards the stereotype threat are stigma consciousness, domain identification, and group identification (Nguyen & Ryan, 2008). According to Brown and Pinel (2003), the performance of individuals who are highly conscious about the negative stereotype ascribed to their social group would be more affected by stereotype threat. Our study did not directly measure the degree of stigma consciousness; however, the score from the gender-stereotype questionnaire showed that our male and female participants had different view of gender stereotype attached to the letter fluency task, even though the effect size was small. Men perceived letter fluency as women-favouring task, while women perceived the task as gender-neutral. Despite being conscious about the stigma, it seems that participants' self-efficacy and educational background protected their performance from the threat.

Studies have shown that the more participants identified themselves with the cognitive domain being evaluated, the more impaired their performance under stereotype threat (Aronson et al., 2002; Keller, 2007; McFarland et al., 2003; Steele & Aronson, 1995). Compared to other gender-sensitive cognitive tests, such as mathematics or mental-rotation tests, letter fluency tests are rarely found on standardized scholastic or admission tests (Bertua et al., 2005). Moreover, the letter fluency test seemed to be very simple in terms of the administration and the responses it required. This might affect how participants assigned importance to their performance. They might perceive this test as less indicative of their verbal ability or even of their intelligence which made them to be less concerned about their performance.

In addition to domain identification, group identification could also affect how participants perform under a gender-stereotype threat. Schmader (2002) found that women with stronger gender identification tend to perform worse on mathematics tests than women with lower gender identification. In our study, participants in the stereotyped groups were asked to indicate their gender to prime their awareness about their gender membership. However, there might be variation in to what extent participants considered this gender membership to be an important part of their selves, and this might affect how our participants reacted to the gender-stereotype threat. Moreover, participants had multiple social identities, and it was possible that participants in this study put more importance to their other group-membership categories than to their gender group-membership (Rydell et al., 2009). As a result, when their gender identity was threatened by the presence of the stereotype, their performance was not affected because they identified less with their gender group.

The results of our study showed

the importance of individuals' personal characteristics (e.g., self-efficacy, educational level, and individuals' susceptibility) to sustaining performance under the interference of the gender-stereotype threat. As far as we know, this is the first study investigating the effect of using various gender stereotype-activation cues to explore how the cues affected performance on letter fluency task. This study also tried to examine how different cognitive loads on letter fluency task interact with the stereotype threat by manipulating the levels of difficulty. Moreover, this study took into account the role of clustering and switching as the underlying components of verbal fluency performance under the stereotype threat. However, our study has several limitations. First, we measured self-efficacy only by a single item self-rating which might not comprehensively represent the construct. Second, we did not measure potential individual moderators of gender stereotype threat. Third, the letter fluency task was not demanding enough for our sample cognitive capacity given their educational background. Fourth, no direct measure of working memory was administered to examine how much an increase in task difficulty corresponded to an increase in cognitive load.

For the reasons explained above, future studies should consider the following: 1) measuring individual psychological constructs that affect the strength of stereotype threat, (e.g., self-efficacy, stigma consciousness, gender identification, domain identification); 2) administering a more complex verbal task, such as the alternate phonemic/semantic fluency test (Costa et al., 2014); it would impose more load on working memory since the rapid shifting between stimuli would require a more effortful and controlled process, and 3) incorporating a direct measure of working memory to enable studying the role of cognitive load on the effect of stereotype threats.

CONCLUSIONS

Our findings contribute to the body of literature on stereotype threat by showing that belief about one's ability, namely, self-efficacy, is an important buffer to the effect of a gender-stereotype threat on both overall performance and underlying cognitive process of the letter fluency task. Other personal characteristics that also need to be considered are educational level, stigma consciousness, domain identification, and group identification. Taken together, our findings suggested the importance of taking into account psychological factors that determine variation in individuals' susceptibility in studying the highly situational phenomenon of stereotype threat.

ACKNOWLEDGEMENTS

REFERENCES

Abwender, D. A., Swan, J. G., Bowerman, J. T., & Connolly, S. W. (2001). Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. *Assessment, 8*(3), 323-338. doi: 10.1177/107319110100800308

Allison, C., Redhead, E. S., & Chan, W. (2017). Interaction of task difficulty and gender stereotype threat with a spatial orientation task in a virtual nested environment. *Learning and Motivation, 57*, 22-35. doi: 10.1016/j. lmot.2017.01.005

Andreoletti, C., & Lachman, M. E. (2004). Susceptibility and resilience to memory aging stereotypes: Education matters more than age. *Experimental Aging Research, 30*(2), 129-148. doi: 10.1080/03610730490274167

Appel, M., Weber, S., & Kronberger, N. (2015). The influence of stereotype threat on immigrants: Review and meta-analysis. *Frontiers in Psychology, 6*, 900. doi: 10.3389/ fpsyg.2015.00900

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*(2), 113-125. doi: 10.1006/ jesp.2001.1491

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology.General, 130*(2), 224-237. doi: 10.1037//0096-3445.130.2.224

Aziz, K. A., Khater, M. S., Emara, T., Tawfik, H. M., Rasheedy, D., Mohammedin, A. S., . . . Qassem, T. (2017). Effects of age, education, and gender on verbal fluency in healthy adult Arabic-speakers in Egypt. *Applied Neuropsychology Adult, 24*(4), 331-341. doi: 10.1080/23279095.2016.1185424

Azuma, T. (2004). Working memory and perseveration in verbal fluency. *Neuropsychology, 18*(1), 69-77. doi: 10.1037/0894-4105.18.1.69

Bajaj, G., Deepa, N., Bhat, J. S., D'Souza, D., & Sheth, P. (2014). Self-efficacy and verbal fluency — does age play a role? *Healthy Aging & Clinical Care in the Elderly,* (6), 17-24. doi: 10.4137/HACCE.S14292

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, USA: W.H. Freeman.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136*(2), 256-276. doi: 2007-06470-007

Bertua, C., Anderson, N., & Salgado, J. F. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology, 78*(3), 387-409. doi: 10.1348/096317905X26994

Borkowski, J. G., Benton, A. L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia, 5*(2), 135-140. doi: 10.1016/0028-3932(67)90015-2

Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology, 39*(6), 626-633. doi: 10.1016/ S0022-1031(03)00039-8

Chung, B. G., Ehrhart, M. G., Ehrhart, K. H., Hattrup, K., & Solamon, J. (2010). Stereotype threat, state anxiety, and specific self-efficacy as predictors of promotion exam performance. *Group & Organization Management, 35*(1), 77-107. doi: 10.1177/1059601109354839

Costa, A., Bagoj, E., Monaco, M., Zabberoni, S., De Rosa, S., Papantonio, A. M., . . . Carlesimo, G. A. (2014). Standardization and normative

data obtained in the Italian population for a new verbal fluency instrument, the phonemic/semantic alternate fluency test. *Neurological Sciences, 35*(3), 365-372. doi: 10.1007/s10072-013-1520-8

Diaz-Asper, C., Schretlen, D. J., & Pearlson, G. D. (2004). How well does IQ predict neuropsychological test performance in normal adults? *Journal of the International Neuropsychological Society, 10*(1), 82-90. doi:10.1017/S1355617704101100

Flore, P. C., Mulder, J., & Wicherts, J. M. (2018). The influence of gender stereotype threat on Mathematics test scores of dutch high school students: A registered report. *Comprehensive Results in Social Psychology, 3*(2), 140-174. doi: 10.1080/23743603.2018.1559647

Ghasemian-Shirvan, E., Shirazi, S. M., Aminikhoo, M., Zareaan, M., & Ekhtiari, H. (2018). Preliminary normative data of Persian phonemic and semantic verbal fluency test. *Iranian Journal of Psychiatry, 13*(4), 288-295.

Hausmann, M., Schoofs, D., Rosenthal, H. E., & Jordan, K. (2009). Interactive effects of sex hormones and gender stereotypes on cognitive sex differences--a psychobiosocial approach. *Psychoneuroendocrinology, 34*(3), 389-401. doi: 10.1016/j.psyneuen.2008.09.019

Hausmann, M. (2014). Arts versus science — academic background implicitly activates gender stereotypes on cognitive abilities with threat raising men's (but lowering women's) performance. *Intelligence, 46*, 235-245. doi: 10.1016/j.intell.2014.07.004

Hendrawan, D. (2013). Factors affecting letter fluency performance among Indonesian undergraduate students: A short report. *Procedia - Social and Behavioral Sciences, 97,* 266-270. doi: 10.1016/j.sbspro.2013.10.232

Hendrawan, D., & Hatta, T. (2010). Evaluation of stimuli for development of the Indonesian version of verbal fluency task using ranking method. *Psychologia, 53*(1), 14-26. doi: 10.2117/psysoc.2010.14

Hendrawan, D., Hatta, T., & Ohira, H. (2015). Do the letters F, A and S represent Indonesian letter fluency stimuli? *Asia-Pacific Psychiatry: Official Journal of the Pacific Rim College of Psychiatrists, 7*(1), 64-71. doi: 10.1111/appy.12082

Hirnstein, M., Andrews, L. C., & Hausmann, M. (2014). Gender-stereotyping and cognitive sex differences in mixed- and same-sex groups. *Archives of Sexual Behavior, 43*(8), 1663-1673. doi: 10.1007/s10508-014-0311-5

Hirnstein, M., Freund, N., & Hausmann, M. (2012). Gender stereotyping enhances verbal fluency performance in men (and women). *Zeitschrift Für Psychologie, 220*(2), 70-77. doi: 10.1027/2151-2604/a000098

Hoffman, B., & Schraw, G. (2009). The influence of self-efficacy and working memory capacity on problem-solving efficiency. *Learning and Individual Differences, 19*(1), 91-100. doi: 10.1016/j.lindif.2008.08.001

Hopko, D. R., Hunt, M. K., & Armento, M. E. (2005). Attentional task aptitude and performance anxiety. *International Journal of Stress Management, 12*, 389-408. doi: 10.1037/1072-5245.12.4.389

Kalpouzos, G., & Eriksson, J. (2013). Memory self-efficacy beliefs modulate brain activity when encoding real-world future intentions. *PlosOne, 8*(9), e73850. doi: 10.1371/journal.pone.0073850

Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat

on female students' maths performance. *The British Journal of Educational Psychology, 77*(Pt 2), 323-338. doi: 10.1348/000709906X113662

Koenig, A. M., & Eagly, A. H. (2005). Stereotype threat in men on a test of social sensitivity. *Sex Roles, 52*(7), 489-496. doi: 10.1007/s11199-005-3714-x

Kosmidis, M. H., Vlahou, C. H., Panagiotaki, P., & Kiosseoglou, G. (2004). The verbal fluency task in the Greek population: Normative data, and clustering and switching strategies. *Journal of the International Neuropsychological Society: JINS, 10*(2), 164-172. doi: 10.1017/S1355617704102014

Lanting, S., Haugrud, N., & Crossley, M. (2009). The effect of age and sex on clustering and switching during speeded verbal fluency tasks. *Journal of the International Neuropsychological Society: JINS, 15*(2), 196-204. doi: 10.1017/S1355617709090237

Mäntylä, T., Ronnlund, M., & Kliegel, M. (2010). Components of executive functioning in metamemory. *Applied Neuropsychology, 17*(4), 289-298. doi: 10.1080/09084282.2010.525090

Mathuranath, P. S., George, A., Cherian, P. J., Alexander, A., Sarma, S. G., & Sarma, P. S. (2003). Effects of age, education and gender on verbal fluency. *Journal of Clinical and Experimental Neuropsychology, 25*(8), 1057-1064. doi: 10.1076/jcen.25.8.1057.16736

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52. doi: 10.1207/S15326985EP3801_6

McFarland, L. A., Lev-Arey, D., & Ziegert, J. C. (2003). An examination of stereotype threat in a motivational context. *Human Performance, 16*(3), 181-205. doi: 10.1207/S15327043HUP1603_2

Moe, A., Hausmann, M., & Hirnstein, M. (2020). Gender stereotypes and incremental beliefs in STEM and non-STEM students in three countries: Relationships with performance in cognitive tasks. *Psychological Research.* doi: 10.1007/s00426-019-01285-0

Mueller, K. D., Koscik, R. L., LaRue, A., Clark, L. R., Hermann, B., Johnson, S. C., & Sager, M. A. (2015). Verbal fluency and early memory decline: Results from the Wisconsin registry for Alzheimer's prevention. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists, 30*(5), 448-457. doi: 10.1093/arclin/acv030

Nguyen, H. H., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *The Journal of Applied Psychology, 93*(6), 1314-1334. doi: 10.1037/a0012702

Pansu, P., Régner, I., Max, S., Colé, P., Nezlek, J. B., & Huguet, P. (2016). A burden for the boys: Evidence of stereotype threat in boys' reading performance. *Journal of Experimental Social Psychology, 65*, 26-30. doi: 10.1016/j.jesp.2016.02.008

Pennington, C. R., Heim, D., Levy, A. R., & Larkin, D. T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *Plos One, 11*(1), e0146487. doi: 10.1371/journal.pone.0146487

Rydell, R. J., McConnell, A. R., & Beilock, S. L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. *Journal of Personality and Social Psychology, 96*(5), 949-966. doi: 10.1037/a0014846

Sanchis-Segura, C., Aguirre, N., Cruz-GÃ³mez, Ã. J., Solozano, N., & Forn, C. (2018). Do gender-related stereotypes affect spatial performance? Exploring when, how and to whom using a

chronometric two-choice mental rotation task. *Frontiers in Psychology, 9*, 1261-1261. doi: 10.3389/fpsyg.2018.01261

Scheuringer, A., & Pletzer, B. (2017). Sex differences and menstrual cycle dependent changes in cognitive strategies during spatial navigation and verbal fluency. *Frontiers in Psychology, 8*, 381. doi: 10.3389/fpsyg.2017.00381

Scheuringer, A., Wittig, R., & Pletzer, B. (2017). Sex differences in verbal fluency: The role of strategies and instructions. *Cognitive Processing,* 1-11. doi: 10.1007/s10339-017-0801-1

Schieman, S. (2001). Age, education, and the sense of control. *Research on Aging, 23*(2), 153-178. doi: 10.1177/0164027501232002

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*(2), 194-201. doi: 10.1006/jesp.2001.1500

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*(2), 336-356. doi: 10.1037/0033-295X.115.2.336

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology, 67*(1), 415-437. doi: 10.1146/annurev-psych-073115-103235

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4-28. doi: 10.1006/jesp.1998.1373

Steele, C. M. (1997). A threat in the air. how stereotypes shape intellectual identity and performance. *The American Psychologist, 52*(6), 613-629. doi: 10.1037/0003-066X.52.6.613

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797-811. doi: 10.1037/0022-3514.69.5.797

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology, 34,* 379-440. doi: 10.1016/S0065-2601(02)80009-0

Steffens, M. C., & Jelenec, P. (2011). Separating implicit gender stereotypes regarding math and language: Implicit ability stereotypes are self-serving for boys and men, but not for girls and women. *Sex Roles, 64*(5), 324-335. doi: 10.1007/s11199-010-9924-x

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: administration, norms, and commentary* (3rd ed.). Oxford, England: Oxford University Press.

Troyer, A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychology, 22*(3), 370-378. doi: 10.1076/1380-3395(200006)22:3;1-V;FT370

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology, 11*(1), 138-146. doi: 10.1037/0894-4105.11.1.138

Van der Elst, W., Van Boxtel, M. P., Van Breukelen, G. J., & Jolles, J. (2006). Normative data for the animal, profession and letter M naming verbal fluency tests for dutch speaking participants and the effects of age, education, and sex. *Journal of the International Neuropsychological Society: JINS, 12*(1), 80-89. doi: S1355617706060115

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39*(5), 456-467. doi: 10.1016/S0022-1031(03)00019-2

Weiss, E. M., Ragland, J. D., Brensinger, C. M., Bilker, W. B., Deisenhammer, E. A., & Delazer, M. (2006). Sex differences in clustering and switching in verbal fluency tasks. *Journal of the International Neuropsychological Society: JINS, 12*(4), 502-509. doi: 10.1017/s1355617706060656

Zimmermann, N., Parente, M. A. d. M. P., Joanette, Y., & Fonseca, R. P. (2014). Unconstrained, phonemic and semantic verbal fluency: Age and education effects, norms and discrepancies. *Psicologia: Reflexão e Crítica, 27*, 55-63. doi: 10.1590/S0102-79722014000100007