

Polish adaptation of emotional Stroop test in assessment of pedophilia – a pilot study

Wojciech Oronowicz-Jaśkowiak¹, Michał Lew-Starowicz^{1,2}

¹ 3rd Department of Psychiatry, Institute of Psychiatry and Neurology

² Department of Psychiatry, Centre of Postgraduate Medical Education

Summary

Aim. A pilot study was conducted in order to construct a Polish adaptation of the emotional Stroop test in assessment of pedophilia.

Material and method. The study consisted of two stages. The first stage involved creating test material by ranking words in adequate lists by competent experts. The second stage consisted of empirical verification of the principle of the emotional Stroop test in a non-clinical population.

Results. Based on the assessment of five competent experts, words were ordered from the most to the least sexually arousing (Kendall's W from 0.368 to 0.693). Six ranked lists were obtained, and the competent experts were subsequently asked to assess whether these lists were suitable for the study (Lawshe's Content Validity Ratio from 0.6 to 1.0). Two categories of words were merged. Five ranked lists were obtained, and the competent experts were subsequently asked again to assess whether these lists were suitable for the study (Lawshe's Content Validity Ratio 1.0). The created lists of words were approved by all competent experts. Based on the experimental study conducted on a non-clinical population, it was shown that, in accordance with the principle of the test, the mean response time for sexually-related words was longer than for neutral words. The mean response time for children-related words did not differ significantly from the response time for neutral words.

Conclusions. Based on the study with competent experts and conducted experiments, an initial Polish adaptation of the emotional Stroop test for diagnosis of pedophilia has been created. Further studies with persons with pedophilia are needed to implement the test in clinical setting.

Keywords: pedophilia, attention control, Stroop task

Introduction

The Stroop task is one of the tools used in cognitive psychology. In the original version, the participant is asked to name as quickly as possible the font colors of the

words displayed on the computer screen. In the compatible condition, a word corresponding to the font color is displayed (e.g., the word “blue” is displayed in blue font), while in the incompatible condition, the font color does not correspond to the word meaning (e.g., the word “red” is displayed in green font). It has been repeatedly shown that in the case of incompatibility of the font color and word meaning, response time is significantly longer [1].

Later, many versions of the modified Stroop test have been presented, which, apart from color names, contained words that could have emotional meaning for certain groups, e.g., alcoholics [2] or persons with depression [3]. The results suggested that persons that had been classified by authors of the study as possibly having an emotional relationship with words referring to specific objects or situations had statistically significant longer response times for naming font colors of such words.

What is important from the point of view of this paper, is that the modified Stroop test has also been tested on sex offenders. Ciardha and Gormley [4] conducted a study on a group of 24 sex offenders against minors, and there was a difference in the processing time for words referring to children when compared with a control group. Price et al. [5] showed that sex offenders display a longer response time for sexually charged words. Similar studies were conducted by Smith and Waterman [6], Price and Hanson [7], and van Leeuwen et al. [8]. However, the interpretation of these results is difficult due to an insufficient number of published papers, lack of homogeneity of study groups, and failure to include a comparison of persons with diagnosed paraphilic disorders and persons without such diagnosis. Moreover, the results of these studies cannot be used in forensic sexology practice, as no normalization adequate for psychological tools has been conducted. It is worth adding, however, that Price and Hanson [7] note that studies leading to creating such tools and their practical applications seem promising. Due to the fact that the potential application of the modified Stroop task could be in the field of forensic sexology, the conclusions of the diagnostic process should be based on particularly thorough studies. Further validation studies are necessary to determine psychometric qualities of the tools.

It seems reasonable to conduct a pilot study aimed at creating lists of Polish words for the emotional Stroop test. In an original study by Smith and Waterman [6], word lists were created based on recommendations by forensic psychologists and prison officers working in therapeutic prison departments for sex offenders. There was no detailed information on the process of creating the word lists, which, from the methodological point of view, should not be acceptable.

In the study by Price and Hanson [7], lists of words were created based on an assessment by five competent experts, who were to decide whether a particular word should be included in the test. However, the competent experts were not sexology experts but students; moreover, for this type of study, it is appropriate to use Lawshe’s content validity ratio (CVR), which the authors do not report to have used. What is more, words were qualified for further studies that, based on the critical value for five competent experts [9], should not have been qualified, as in the case of a disagreement

between two experts out of five the CVR is lower than the CVR threshold = 1.00. Furthermore, Price and Hanson [7] did not pre-rank the generated lists based on the level of arousal in persons with pedophilic preferences. It was only indicated whether a particular word should be qualified. In the study by Smith and Waterman [6], there is a lack of information concerning the method used for selecting words (ranking or categorical selection). In view of the above, it seems therefore necessary to create new lists, as the available research contains significant methodological errors. A tool for clinical use should meet high methodological criteria, including a strict construction process of the tool.

It is also justified to verify whether the Polish adaptation of the emotional Stroop task corresponds to the results of studies conducted in English, in particular when using vulgar and non-vulgar sexual words. A limitation in the interpretation of results is a lack of division between vulgar and non-vulgar sexual words. It is concluded that sexual words should be characterized by longer response times, but it is difficult to determine whether the reason for the longer response time is the unexpected occurrence of a vulgar word or greater emotional meaning of the sexual word for the participant. For this reason, it seems important to distinguish between these groups of words.

Hypotheses

Stage 1 of the tool construction

Due to the exploratory nature of the study with competent experts, no research hypotheses were made. The aim of the study was to obtain lists of ranked words that would contain words from least to most arousing for pedophiles, as perceived by competent experts.

Stage 2 of the tool construction

If the lists of words have been correctly adapted to Polish conditions, similar results should be expected in terms of adequate differences in response time between different word categories. Therefore, if it is reported that sex offenders against minors present a longer response time for words referring to children [4, 7], then it should be expected that persons from a non-clinical population will not present this pattern. Similarly, if it is reported that persons who self-identify as pedophiles present a longer response time for words referring to children [4], then it also should be expected that persons from a non-clinical population will not present this pattern. Otherwise, the relevance of the Stroop test for assessment of pedophilic preferences could be questioned.

Based on the presented literature review, research hypotheses regarding adjusted mean response times for words from different categories of the emotional Stroop test were set. It was expected that the adjusted mean response time to determine font color for words denoting color names (H1) and vulgar words (H2) will be longer when compared to the response time for words related to children.

Material and method

Stage 1 of the tool construction

In order to determine the lists of words that the experts consider the best stimuli for the Polish adaptation of the emotional Stroop test, a study with sexology experts was conducted. The selection process of the final lists of words consisted of three sub-stages: (1) creation of initial word lists; (2) expert assessment through ranking the words; and (3) re-evaluation by the experts who were to accept or reject the final word lists.

The first sub-stage of the study consisted in generating lists of words that were subsequently submitted for evaluation by the competent experts. These lists were created based on stimuli words translated into Polish by Kasperek (2015) that were previously used in the study by Price and Hanson [7], and additionally supplemented with words found in a thesaurus.

The second sub-stage consisted in ranking the individual words. The experts were to rank the words from most to least arousing for persons with pedophilic preferences, separately for each group of words. The participants were five competent sexology experts, including: two psychologists – clinical sexologists certified by the Polish Sexological Society (PTS – Pol. *Polskie Towarzystwo Seksuologiczne*); one medical doctor specialized in sexology and clinical sexology certified by the PTS; and two psychologists who had completed post-graduate clinical sexology studies and were in the process of PTS certification. The experts were personally invited to participate in the study.

The third sub-stage consisted in an assessment whether the obtained word lists could be used for future research. The experts for this stage were the same persons as in sub-stage two. The task of these competent experts was to make a categorical decision regarding whether the word lists created in sub-stage two would be appropriate for the study. Similarly as in the second sub-stage, the experts evaluated each list separately.

Stage 2 of the tool construction

In order to determine the differences between adjusted mean response times, an empirical study was conducted, in which the participants had to perform two tasks containing different sets of words, each comprising the same total number of words (Table 1).

Table 1. Lists of words presented to the participants for the two tasks

Task 1	Task 2
"Neutral" list of words (n = 25)	"Neutral" list of words (n = 25)
"Colored" list of words (n = 25)	"Colored" list of words (n = 25)
"Pedophile – new" list of words (n = 15)	"Pedophile – new" list of words (n = 15)
"General sexual – vulgar" list of words (n = 15)	"General sexual – mild" list of words (n = 15)
"Pedophile – boys" list of words (n = 15)	"Pedophile – girls" list of words (n = 15)

The study group consisted of 25 persons ($M = 37.25$ years; $SD = 15.56$), including 16 women ($M = 39.33$ years; $SD = 14.42$) and 9 men ($M = 33.78$ years; $SD = 16.72$). A non-random ad-hoc sampling strategy was employed. In the study group, 22 persons (88%) had higher education, and 3 persons (12%) had secondary education.

The study was conducted on a notebook computer. Study instructions and the test were displayed on the computer screen. The participant was to press the key corresponding to the font color of the displayed word. Study instructions were the same for both tasks:

“During the following tests, you will see words displayed in different font colors. Your task is to indicate the font color of the word, ignoring the actual meaning of the word. Indicate the color of the word by pressing the appropriate key: d for red font, f for green font, j for blue font, k for yellow font. Try to answer as quickly and accurately as possible, as your response time will be measured. If you make a mistake, a red X will be displayed on the screen. Press continue when you are ready to begin the training session.”

Before starting the actual test, the participant was asked to complete a training session by pressing the appropriate key corresponding to the font color of 10 numerals. The participant determined the font color by pressing the appropriate key on the computer keyboard. The results of this part of the test were not recorded. The aim of the training session was to familiarize the participants with the study procedure and to practice.

After the training session was completed, the test automatically began. The participants were presented with all the words included in each category and were asked to determine the font color of each word as fast as possible. The words were displayed in random order, and font colors were also assigned randomly. The list of “neutral” words included words devoid of emotional meaning, while the list of “colored” words included names of colors (see Annex 1). The list of words “general sexual – vulgar” (see Table 3) and “general sexual – mild” (see Table 2) included words related to human sexuality. The lists of words “pedophile – new” (see Table 10), “pedophile – boys” (see Table 6) and “pedophile – girls” (see Table 5) included words related to children and childhood. Each list contained 15 words that were obtained with the participation of the competent experts.

The response times for each word and number of errors were automatically recorded by the psychological measurement software – Inquisit.

Results

Stage 1 of the tool construction

Based on the presented sources, a list of 180 words was created and each word was assigned to one of six categories, each category containing 30 words. Individual words were categorized based on their subject – the lists of words included: words concerning sexual relationships (list 1 and 2), children and childhood (list 3 and 6), girls (list 4), and boys (list 5).

In order to determine the agreement among the competent experts, Kendall's W coefficient of concordance for each category was calculated. The ranked words are presented in Tables 2-7. Table 8 presents Kendall's W for each category of words. Subsequently, in order to determine the level of agreement among the experts, Lawshe's content validity ratio (CVR) was calculated for each category. The results are presented in Table 9.

Table 2. List of words in the “general sexual – mild” category ordered from lowest to highest rank

List 1					
Item	Word	Mean rank*	Item	Word	Mean rank*
1	wet [Polish synonym 1]	6	16	touch	16.2
2	caress	7.2	17	vagina	16.2
3	undress	8	18	panties	16.4
4	sex	8.8	19	striptease	16.8
5	naked	9	20	massage	17.8
6	tempt	9.2	21	vibrator	18
7	moan	9.4	22	thongs	18
8	suck	9.6	23	flirt	18.8
9	orgasm	10	24	lover	19.8
10	buttocks	10.4	25	masturbation	22.8
11	arouse	11.6	26	erection	22.8
12	breasts	13.2	27	anus	23.4
13	virgin	14.8	28	prostitute	24.4
14	lick	15	29	love	26.2
15	kiss	16	30	rub	29.2

* The lower the rank, the more arousing the word was considered to be by the experts, as the experts ranked all 30 words from 1 to 30, where 1 corresponded to the most arousing word, and 30 – to the least arousing.

Table 3. List of words in the “general sexual – vulgar” category ordered from lowest to highest rank

List 2					
Item	Word	Mean rank	Item	Word	Mean rank
1	cunt	4	16	fuck [Polish synonym 1]	17.4
2	oral	4.2	17	saliva	17.8
3	anal	4.8	18	fuck [Polish synonym 2]	18
4	wet [Polish synonym 2]	5.6	19	take	18.2
5	boobs	6.6	20	whore	18.2
6	ass	6.8	21	hard	18.4
7	fuck [Polish synonym 3]	7.6	22	hot	19.4
8	fuck [Polish synonym 4]	8	23	fuck [Polish synonym 5]	20.6
9	slit	8.4	24	hair	22.4
10	fuck [Polish synonym 6]	12.2	25	fuck [Polish synonym 7]	22.4
11	shoot a load	13.4	26	fuck [Polish synonym 8]	23.8
12	fuck [Polish synonym 9]	14.2	27	buckle	24
13	fuck [Polish synonym 10]	14.8	28	fuck [Polish synonym 11]	26.6
14	dick	15.2	29	fuck [Polish synonym 12]	27
15	fuck [Polish synonym 13]	16.4	30	fuck [Polish synonym 14]	28.6

Table 4. List of words in the “pedophile – mild” category ordered from lowest to highest rank

List 3					
Item	Word	Mean rank	Item	Word	Mean rank
1	girl	4.6	16	mouth	16.2
2	boy	5.6	17	diapers	17.2
3	bottom	5.8	18	kitty	17.8
4	children	7	19	childhood	18
5	play	9	20	merry-go-round	19.2
6	secret [Polish synonym 1]	9.2	21	swing	19.6

table continued on the next page

7	pussy [Polish synonym 1]	9.6	22	swaddle	19.6
8	toys	12.2	23	ducky	19.8
9	innocence	12.8	24	bib	20
10	secret [Polish synonym 2]	12.9	25	crayons	21.2
11	kitten	13.4	26	rompers	21.6
12	kindergarten	13.8	27	dependence	21.6
13	primary school	14	28	classes	22.2
14	school	15.4	29	care	24.8
15	pacifier	15.6	30	tenderness	25.3

Table 5. List of words in the “pedophile – girls” category ordered from lowest to highest rank

List 4					
Item	Word	Mean rank	Item	Word	Mean rank
1	gap	4.4	16	punk	15
2	minnie	5	17	snot girl [Polish synonym 1]	15.4
3	doll	7	18	tweenie	16.8
4	fanny	7.4	19	immature	17.2
5	sweet	9.8	20	sissy	17.2
6	girl	11	21	snot girl [Polish synonym 2]	17.6
7	pussy [Polish synonym 2]	11.2	22	dependence	19.2
8	miss [Polish synonym 1]	11.4	23	teddy girl	19.9
9	miss [Polish synonym 2]	13.6	24	teenager	19.9
10	pigtails	13.6	25	young	20.8
11	mischief	13.8	26	dress	21.2
12	snot girl [Polish synonym 3]	14.4	27	toddler	21.8
13	innocent	14.4	28	maiden	23.2
14	daughter	14.8	29	juvenile	24.4
15	princess	14.8	30	cream	28.8

Table 6. List of words in the “pedophile – boys” category ordered from lowest to highest rank

List 5					
Item	Word	Mean rank	Item	Word	Mean rank
1	boy	5.6	16	colt	16.4
2	little fellow	7	17	son	16.8
3	snot boy [Polish synonym 1]	7.2	18	tin soldiers	17.2
4	wiener	7.6	19	young lad	17.4
5	snot boy [Polish synonym 2]	8	20	juvenile	17.8
6	baby [Polish synonym 1]	10.4	21	immature	18.6
7	sissy	11.4	22	toddler	19.2
8	dependent	11.6	23	young man	20
9	innocent	12.2	24	youngster	20.4
10	snot boy [Polish synonym 3]	12.4	25	young	21
11	boy [Polish synonym 3]	12.6	26	teenager	21.2
12	punk	14.4	27	sweet	21.8
13	rascal	14.8	28	adolescent	22.6
14	kid [Polish synonym 1]	15.6	29	underage	23.4
15	sonny	15.8	30	caring	24.6

Table 7. List of words in the “pedophile – general” category ordered from lowest to highest rank

List 6					
Item	Word	Mean rank	Item	Word	Mean rank
1	bottom	5.8	16	childhood	14.8
2	children	6.2	17	bimbo	15.8
3	preschooler	8.8	18	baby [Polish synonym 2]	16.8
4	little hands	9.2	19	rompers	17.4
5	secret [Polish synonym 1]	10	20	lull	17.6
6	baby [Polish synonym 3]	10.2	21	play	18
7	on my lap	10.4	22	school	18
8	kid [Polish synonym 2]	12.2	23	pacifier	19
9	toys	12.6	24	diapers	19
10	play	13	25	milk	19.4
11	hug	13	26	classes	22

table continued on the next page

12	kindergarten	13.4	27	care	22.8
13	baby [Polish synonym 4]	14.2	28	crayons	24
14	daddy	14.2	29	bib	24.4
15	primary school	14.4	30	skin	28.4

Table 8. **Kendall's W for each list of words**

No. of list	Test statistics
1	Kendall's W = 0.486; chi square = 70.517; df = 29; p < 0.001
2	Kendall's W = 0.693; chi square = 100.489; df = 29; p < 0.001
3	Kendall's W = 0.431; chi square = 62.565; df = 29; p < 0.001
4	Kendall's W = 0.424; chi square = 61.492; df = 29; p < 0.001
5	Kendall's W = 0.368; chi square = 53.320; df = 29; p = 0.004
6	Kendall's W = 0.385; chi square = 55.885; df = 29; p = 0.002

Table 9. **Lawshe's content validity ratio for each list of words**

	List 1	List 2	List 3	List 4	List 5	List 6
No. of answers	5	5	5	5	5	5
No. of positive answers	5	5	4	5	5	4
No. of negative answers	0	0	1	0	0	1
CVR	1.00	1.00	0.60	1.00	1.00	0.60

Lawshe's content validity ratio was 1.00 for four categories of words and 0.6 for two categories, exceeding the critical value allowing a given assessment to be acceptable for five experts [9]. Four experts considered the created word lists appropriate for further empirical studies. One expert found that the categories of words "pedophile – general" and "pedophile – mild" should not be included in further studies, as they cover a similar thematic scope; moreover, there are recurrent words between the lists. As suggested by the expert, the categories "pedophile – general" and "pedophile – mild" have been merged, creating a new word list – "pedophile – new" (list 6b). This list included 15 words rated as the most arousing from both merged categories (see Table 10).

Table 10. **Merged list of words included in the final list, ordered from the lowest to highest rank**

List 6b					
Item	Word	Mean rank	Item	Word	Mean rank
1	girl	5.8	31	pacifier	15.6

table continued on the next page

2	boy	6.2	32	bimbo	15.8
3	bottom*	8.8	33	mouth	16.2
4	bottom	9.2	34	baby [Polish synonym 2]	16.8
5	children*	10	35	diapers	17.2
6	children	10.2	36	rompers	17.4
7	preschooler*	10.4	37	lull	17.6
8	play*	12.2	38	kitty	17.8
9	secret [Polish synonym 1]*	12.6	39	childhood	18
10	little hands*	13	40	play	18
11	pussy [Polish synonym 1]*	13	41	school	18
12	secret [Polish synonym 1]	13.4	42	pacifier	19
13	baby [Polish synonym 3]*	14.2	43	diapers	19
14	on my lap*	14.2	44	merry-go-round	19.2
15	toys*	14.4	45	milk	19.4
16	kid [Polish synonym 2]*	12.2	46	swing	19.6
17	toys	12.6	47	swaddle	19.6
18	innocence*	12.8	48	ducky	19.8
19	secret [Polish synonym 2]*	12.9	49	bib	20
20	play	13	50	crayons	21.2
21	hug*	13	51	rompers	21.6
22	kitten*	13.4	52	dependence	21.6
23	kindergarten	13.4	53	classes	22
24	kindergarten	13.8	54	classes	22.2
25	primary school	14	55	care	22.8
26	baby [Polish synonym 4]	14.2	56	crayons	24
27	daddy	14.2	57	bib	24.4
28	primary school	14.4	58	care	24.8
29	childhood	14.8	59	tenderness	25.3
30	school	15.4	60	skin	28.4

* 15 most arousing words, excluding repetitions

After creating the new list, all the lists were presented to the experts for re-evaluation. In order to determine agreement between the experts, Lawshe's content validity ratio was recalculated for each category. The results are presented in Table 11.

Table 11. Lawshe's content validity ratio for each list of words

	List 1	List 2	List 3	List 4	List 5	List 6b
No. of answers	5	5	5	5	5	5
No. of positive answers	5	5	5	5	5	5
No. of negative answers	0	0	0	0	0	0
CVR	1.00	1.00	1.00	1.00	1.00	1.00

Stage 2 of the tool construction

Following the research by Price et al. [5], before calculations were made, the response time was adjusted by subtracting the average response time for each category (“colored”, “pedophile – new”, “general sexual – vulgar”, “pedophile – boys”, “general sexual – mild”, “pedophile – girls”) from the average response time for neutral words.

Task 1

Analysis of variance (ANOVA) was conducted in order to determine whether there were differences between the adjusted mean response time for different categories under the appropriate study conditions. Based on Mauchly's sphericity test, it was concluded that the assumption of sphericity was not met (Mauchly's $W = 0.528$; $p = 0.013$); hence, the Greenhouse-Geisser correction method was applied. Statistically significant differences were observed ($F = 4.249$; $p = 0.018$); therefore, post-hoc LSD tests were conducted. It has been demonstrated that the mean adjusted response time for words from the “vulgar” category was longer than for words from the “pedophile – boys” category (difference of means = 10740.12 ms; $p = 0.033$; 95% CI (956.67 ms, 20503.57 ms)). The mean adjusted response time for words from the “colored” category was longer than for the “pedophile – new” category (difference of means = 6823.40 ms; $p = 0.044$; 95% CI (185.40 ms, 13461.40 ms)) and longer than for the “pedophile – boys” category (difference of means = 11384.60 ms; $p = 0.006$; 95% CI (3580.24 ms, 19188.96 ms)).

Task 2

Analysis of variance (ANOVA) was conducted afresh in order to determine whether there were differences between the adjusted mean response time for different word categories under the appropriate study conditions. Due to the lack of statistically significant differences ($p > 0.05$), no further statistical analyses were conducted.

Discussion

Stage 1 of the tool construction

The study consisted in ranking words from most to least sexually arousing by competent sexology experts. The aim of the study was to create appropriate categorized lists of words for the Polish adaptation of the Stroop test for the assessment of pedophilic preferences. Initially, a list of 180 words divided into 6 categories was created.

The obtained lists were to be used as a starting point for creating lists of words that could be used for further study. The limitations of the study include, however, low concordance coefficients of the experts on the ranked lists of words during the second stage. The concordance coefficients were low for all the lists, excluding the “general sexual” list of words. Due to the low sensitivity of the ranked words, it was impossible to accurately deduce the rank differences between individual words. Hence, in further studies it was planned to use only half of the highest-rated words from each of the five lists.

Other limitations of the study include the fact that the experts who assessed the words as arousing for persons with pedophilic preferences were sexology experts, and were not themselves pedophiles. It seems that an assessment of words conducted by pedophiles could be more accurate, as it would not be based on theoretical knowledge on pedophilia, but real-life experience. However, it should be noted that recruiting pedophiles for this kind of study could be considered unethical, and their assessment could be falsified, especially in the case of participants currently undergoing sex offender therapy.

Stage 2 of the tool construction

In this stage of the study, the differences between mean adjusted response times were calculated. As expected, it was found that participants of the study from the non-clinical population took longer to determine the font color of words denoting color names (H1), as well as words with emotional connotations, i.e., vulgarisms (H2), when compared with the response time for words related to children and childhood.

The obtained results are consistent with previous reports, according to which persons from the control group have a shorter response time for children-related words when compared with persons serving sentences for sexual assault of minors; however, the authors of these previous studies did not provide sufficient information on the diagnosis of disorders of sexual preference [4, 7].

In this study, persons from the non-clinical population, assumed not to have pedophilic preferences, had shorter response times for words related to childhood compared to the response time for words with emotional meaning in the first task. What is interesting is that this effect was not observed for Task 2, which was performed directly after Task 1. These results suggest that completing the emotional Stroop test for the first time significantly affects the performance during repeated testing. However, it

is impossible to determine whether the lack of this effect is a result of presenting the word list “general sexual – mild” second, or whether it is a result of the participants’ lack of emotional response to the presented words.

At least two conclusions can be drawn from these observations. Firstly, in future research studies, it is not advisable to test the participants twice. Secondly, in order to minimize the habituation effect, it may be necessary to reduce the individual lists of words.

The limitations of this study include a small sample group, unequal gender representation, and a predominance of participants with higher education.

General discussion

Based on our research, we conclude that future empirical studies using the word lists created by our team should only include half of the words from each category. The above recommendation was formulated on the basis of two arguments.

First of all, further studies should include words ranked as most sexually arousing from each category. We do not recommend using the more extensive list of words, as low expert concordance coefficients do not allow for sensitive differentiation between individual words. After reducing the word lists by half, all the lists were accepted by the experts.

Second of all, based on the empirical studies conducted on the non-clinical population, it can be assumed that the results are strongly affected by the habituation effect. It seems that presenting too many stimuli reduces the participant’s emotional response to the presented words.

The presented lists of words and current research guidelines do not support the use of the Polish version of the emotional Stroop task for diagnostic purposes. There are two rationales for which the presented lists cannot currently be used for sexological diagnosis. Firstly, further studies are needed to verify whether the created lists differentiate persons with a history of sexual assault of minors and with pedophilic preferences from persons from the non-clinical population. Secondly, it is necessary to develop psychometric indicators and create appropriate normalization data for individual groups of sex offenders.

Annex 1

“Neutral” and “colored” list of words

“Neutral” list of words				“Colored” list of words			
Item	Word	Item	Word	Item	Word	Item	Word
1	board	16	box	1	purple	16	golden
2	car	17	map	2	gray	17	silver
3	motorcycle	18	rope	3	pale	18	jade
4	photos	19	washing machine	4	pink	19	topaz
5	microwave	20	printer	5	orange	20	emerald
6	toaster	21	scanner	6	scarlet	21	purple
7	kettle	22	album	7	maroon	22	gray
8	glasses	23	intercom	8	crimson	23	pale
9	binder	24	vacuum cleaner	9	white	24	pink
10	computer	25	cable	10	black	25	orange
11	hard drive			11	cyan		
12	scarf			12	brown		
13	stone			13	tin		
14	projector			14	bronze		
15	ring			15	violet		

References

1. Nęcka E, Orzechowski J, Szymura B. *Psychologia poznawcza*. Warszawa: Wydawnictwo Szkoły Wyższej Psychologii Społecznej; 2008.
2. Christiansen P, Bloor JF. *Individualised but not general alcohol Stroop predicts alcohol use*. *Drug Alcohol Depend*. 2014; 134: 410–413.
3. Hill AB, Knowles TH. *Depression and the ‘emotional’ Stroop effect*. *Pers. Individ. Differ*. 1991; 12(5): 481–485.
4. Ciardha CÓ, Gormley M. *Using a pictorial-modified Stroop task to explore the sexual interests of sexual offenders against children*. *Sex. Abuse* 2012; 24(2): 175–197.
5. Price SA, Beech AR, Mitchell I, Humphreys GW. *Measuring deviant sexual interest in adolescents using the emotional Stroop task*. *Sex. Abuse* 2014; 26(5): 450–471.
6. Smith P, Waterman M. *Processing bias for sexual material: The emotional Stroop and sexual offenders*. *Sex. Abuse* 2004; 16(2): 163–171.

7. Price SA, Hanson KR. *A modified Stroop task with sexual offenders: Replication of a study*. J. Sex. Aggress. 2007; 13(3): 203–216.
8. Leeuwen van ML, Baaren van RB, Chakhssi F, Loonen MG, Lippman M, Dijksterhuis A. *Assessment of implicit sexual associations in non-incarcerated pedophiles*. Arch. Sex. Behav. 2013; 42(8): 1501–1507.
9. Ayre C, Scally AJ. *Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation*. Meas. Eval. Couns. Dev. 2014; 47(1): 79–86.

Address: Wojciech Oronowicz-Jaśkowiak
Institute of Psychiatry and Neurology
3rd Department of Psychiatry
02-957 Warszawa, Sobieskiego Street 9
e-mail: oronowicz@icloud.com