# Monitoring cognition before, during and after awake brain tumor surgery; reliability of a cognitive screener

Master Thesis Neuropsychology

Utrecht University

June 2020

**Name:** I.B. Lanenga

**Student number:** 5682363

**Supervisor:** dr. C. Ruis

# Index

**Abstract** ..... **3**

**Introduction** ..... **4**

**Methods** ..... **7**

***Participants*** ..... ***7***

***Procedure*** ..... ***7***

***Material*** ..... ***7***

*Snodgrass Naming Task* ..... *8*

*Reading* ..... *8*

*Stroop Test with and without block* ..... *8*

*Digit Span Forward and Backward* ..... *9*

*Dot Counting Test with background* ..... *9*

*How are They Feeling?: Colorcards* ..... *9*

***Statistical Analyses*** ..... ***10***

**Results** ..... **11**

***Demographic and clinical characteristics*** ..... ***11***

***Neuropsychological tests*** ..... ***12***

1. *Snodgrass Naming Task* ..... *12*
2. *Stroop Test without block and Stroop Test with block* ..... *12*
3. *Digit Span Forward and Backward* ..... *13*
4. *Reading* ..... *14*
5. *Dot Counting Test with background* ..... *15*
6. *How are They Feeling?: Colorcards* ..... *15*

**Discussion** ..... **17**

**References** ..... **22**

**Appendix** ..... **28**

***I: Answer Sheet Snodgrass Naming Task*** ..... ***28***

***II: Answer Sheet Stroop Task without Block*** ..... ***29***

***III: Answer Sheet Stroop Task with Block*** ..... ***30***

## Abstract

**Background:** Cognitive assessment during awake brain tumor surgery is used to maximize the extent of tumor resection while minimizing the risk of cognitive damage in patients. Assessments of language functions are widely reported while other cognitive domains are underexposed. In this study, the current cognitive screener that is used at the UMC Utrecht Department of Neurosurgery is expanded to monitor a broader spectrum of cognitive domains. Furthermore, parallel versions have been made to assess cognitive functioning over time.

**Methods:** Four cognitive screeners were administrated to healthy Dutch individuals (N=38). Each version tapped different domains: language (object naming, reading), executive functioning and attention (Stroop Test with and without Block), working memory (Digit Span Forward and Backwards), visual perception (Dot Counting Test with Background) and emotion recognition (How are They Feeling?: Colorcards). One way ANOVA and Kruskal-Wallis analysis were performed to determine significant group differences.

**Results:** No significant differences between groups in the Stroop Test without Block ($p$=.025), Digit Span Forward ($p$=.531) and Backwards ($p$=.079), Reading (words/sentences) and Dot Counting Test were found. Significant differences between groups were found for the Snodgrass Naming Task between two versions ($p$=.017) and How are They Feeling?: Colorcards ($p$=.000).

**Conclusion:** This study demonstrated the possibility to expand the cognitive screener and conducting reliable parallel versions to monitor cognitions in multiple timeframes, considering no significant differences between versions have been found in the majority of neuropsychological tests. This provide insight in which domains extensive test assessments is needed for optimal patient care. Investigations into an emotion recognition task is desired.


*Keywords:* awake brain tumor surgery, monitoring, cognitive functioning, neuropsychological assessment, test assessment

## Introduction

Over the last decades, neurosurgical interventions in brain tumors have been reported more often (Duffau, 2010). Although most neurosurgical interventions are still performed under general anesthesia, a shift has been made to where surgical neuro-oncology interventions are performed entirely under local anesthesia, named 'awake-brain surgery' (Whittle, Midgley, Georges, Pringle & Taylor, 2005; Dziedzic & Bernstein, 2014). The main purpose of awake brain tumor surgery is to maximize the extent of abnormal brain resection and to minimize the risk of post-operative damage in patients. This is the so called 'onco-functional balance', which is defined by the compromise between achieving the maximum tumor resection together with preservation of the maximum brain functions in patients (Coello et al., 2013; Duffau & Mandonnet, 2013).

Furthermore, general anesthesia could be avoided, which reduces the need for post-operative intensive care monitoring of patients and results in a shorter length of stay in the hospital (Whittle, Midgley, Georges, Pringle & Taylor, 2005). Besides, research shows that the morbidity rate in neuro-oncology patients who underwent awake brain surgery is significant lower compared to patients who underwent general anesthesia (Sanai, Mirzadeh & Berger, 2008; Conte et al., 2008; Freyschlag & Duffau, 2014).

Intra-operative monitoring of brain functions during awake brain surgery has been done by awake brain mapping, which is a method where the neurosurgeon stimulates eloquent areas around tumor to precisely locate functional areas of the brain (Freyschlag & Duffau, 2014). The main goal of awake brain mapping is to optimize safe tumor removal and it is proven to be the most reliable method in awake-brain surgery to optimize the onco-functional balance (Duffau & Mandonnet, 2013). Awake brain tumor surgery allows intra-operative patient assessment through monitoring brain functions, which is done by administering neuropsychological tests. Monitoring cognitive functions during awake brain surgery is not comparable to standard neuropsychological assessment. Whereas specificity is important during standard neuropsychological test assessment, sensitivity is most essential in intra-operative monitoring of cognition for noticing subtle changes in cognition (Ruis, 2018). Sensitivity refers to the probability to detect cognitive deficits and is defined as the true positive rate, whereas specificity refers to the probability to detect no cognitive deficits of individuals who do not meet diagnostic criteria for the certain deficit and is defined as the true negative rate (de Jager, Hogervorst, Combrinck, Budge, 2003). Besides the importance of sensitivity, intra-operative neuropsychological assessments should also be short in presentation of stimuli and response,

have a low chance level and learning effects should be minimalized (Coello et al., 2013; Van Zandvoort, Ruis & Hendriks, 2016).

Surprisingly, literature shows that the majority of awake brain surgery interventions have been performed almost exclusively in cases of lesions that involve language areas (Sanai, Mirzadeh & Berger, 2008; Santini et al., 2012). Also, assessments for motor functions are frequently reported, but are not monitored by a specific task but mostly tested through performing movements or describing sensations (Sala & Lanteri, 2003; Ruis, 2018). Nevertheless, reported post-operative cognitive deficits in patients that might occur are not exclusively in the language domain. Visuospatial, memory, attention, emotional and behavioral deficits have regularly been observed after brain surgery but are mainly present in memory and attention (Yoshii et al., 2008; Duffau, 2010; Santini et al., 2012). Intra-operative monitoring of other cognitive domains such as memory, emotion recognition and visuospatial are currently underexposed (Ruis, 2018). Nevertheless, monitoring non-language functions during awake brain surgery is required to preserve the maximum of cognitive functions to optimize safe tumor resection. As mentioned before, given the specific criteria that intra-operative tests or paradigms must meet, not all available reliable and valid test assessments can be implemented. Therefore, to achieve an optimal onco-functional balance in patients who underwent open brain tumor surgery, there is a need for improvements or modifications of existing standard neuropsychological test or paradigms (Ruis, 2018).

Although reliable cognitive screeners are widely available, such as the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) and the Mini Mental State Examination (MMSE; Cockrell & Folstein, 2002), these screeners are not feasible for intra-operative screening. Therefore, the UMC Utrecht Department of Neurosurgery has compiled and implemented a cognitive screener for monitoring a broad spectrum of cognitive functions during awake brain surgery over the past 10 years. This test set is used pre-operative, to indicate the baseline measurement of cognitive functions and to practice so that the patient knows what to expect during surgery. Furthermore, the same test is used intra-operative just before the tumor resection to get an indication of the actual baseline level of multiple cognitions. During the brain mapping and tumor resection the same tests that are part of the cognitive screener are administrated, to notice possible changes in cognitions compared with patients' pre-operative results and impressions. In contrast to other studies and institutions where awake brain surgery is reported or performed, the cognitive screener that has been used in the UMC Utrecht does not solely focus on language and motor functions. Non-language domains such as executive functioning and visuoperception are also mapped by this test set and enables screening of

various cognitive functions in a short period of time during awake brain surgery in language and non-language areas.

Despite the fact that the current cognitive screener has been in use successfully over the past decade, some improvements are desirable to optimize mapping of cognitive functions pre- intra- and post-operative. First, the current cognitive screener lacks monitoring emotion recognition, which can be important in patients with frontal lobe lesions (Andrewes et al., 2003; Heberlein, Padon, Gillihan, Farah & Fellows, 2008). Besides that, measurements for the language domain can be expanded to get an even more specific representation of this function. Therefore, optimizing the current cognitive screener by expanding the test set is desirable to map an even broader spectrum of cognitive domains in a more sensitive and reliable way. When expanding the current cognitive screener, it is crucial to keep in mind that this test set must be suitable for a broad population varying in age and educational level and be feasible during surgery.

An expanded cognitive screener can offer a step in the future to measure cognitions in a standardized way at different moments in time. In order to achieve this, developing several parallel versions is necessary for reliable measurements and is required to minimize the possible occurrence of a practice effect (Benedict & Zgaljardic, 1998; Duff et al., 2007). Various factors must be taken into account when constructing several parallel versions. The tests must be the same in terms of difficulty and administration time and must have equivalent and not identical stimuli. Constructing parallel versions offers the possibility to map cognitions not only pre- and intra-operative but also in different time frames post-operative, for example three months after surgery. This provides long term information about the course of cognitive functioning after awake brain surgery.

In short, monitoring cognition pre-operative and during awake brain surgery is done by a cognitive screener that consist of several short neuropsychological tests. Nevertheless, there is a need to expand the current test set that has been used to measure a broader spectrum of neuropsychological cognitions and functions in a reliable, sensitive way. This study aimed to investigate whether it is possible to expand the current cognitive screener that is used at the UMCU and conduct reliable parallel versions of this cognitive screener to monitor cognition over time. It is important to keep in mind the applicability and requirements for use during awake brain tumor surgery, where test material should be sensitive and short in presentation and response time. We hypothesize that each neuropsychological test in each version will give reliable results in the tested cognitive domains and functions when administered to healthy subjects. Therefore, each cognitive screener will be reliable to be implemented during awake

brain tumor surgery and be suitable and sensitive enough to detect subtle changes in cognitive functioning in patients.

## Methods

### *Participants*

A total of 38 healthy Dutch participants, recruited by convenience sampling from the broad network of the researcher and throughout poster advertisement, volunteered to participate in this study. Inclusion criteria were (a) no history of presence of neurological disorder or brain injury; (b) presence of psychiatric disorder; (c) no history or current alcohol or drugs abuse. The participants had to have a minimum age of 18 years to participate in this study. No maximum age was determined. To classify the level of education, the Dutch Verhage scale was used (Verhage; 1983) (1=less than primary school, 7=university). The seven categories were merged into three ordinal categories: low educational level (Verhage 1,2,3,4), middle educational level (Verhage 5) or high educational level (Verhage 6,7).

### *Procedure*

All participants provided written informed consent and filled in a demographic information questionnaire where gender, age, native language, educational level, psychiatric history/existence and neurological history/disorders (dyslexia, ADHD, ASD, color-blindness) were examined. All participants performed the four tests in random order, to eliminate any possible response biases. The tests were, depending on participants' preference, administered individually at participants' home or at researchers' home. The cognitive screeners were presented on a laptop and presented in Microsoft Office PowerPoint software. Each test took approximately five minutes. The total time needed to complete the total test was approximately 30 minutes. Tests were taken immediately following each other, without breaks. Raw scores were written down on an answer sheet by researcher and differ in each subtest.

### *Material*

In order to assess a broad spectrum of cognitive functions, six neuropsychological tests to screen different cognitive domains are presented in the same order in four versions with different stimuli. Pictures from the Snodgrass Naming Task (Snodgrass & Vanderwart, 1980)

for object naming and reading out loud words, sentences and two short stories were conducted to cover the language domain. The Stroop Test (with and without block), based on the D-KEFS version (Delis, Kaplan & Kramer, 2001) was used to cover the attention and executive functioning domains. For memory, the Digit Span Backward and Forward were used (Wechsler, 2008). For visual perception, the Dot Counting Test with Background was administered (James, Plant & Warrington, 2001). Finally, How are they Feeling?: Colorcards (Speechmark Publishing Ltd, 2012) were used for emotion recognition. An overview of the assessed neuropsychological tests and corresponding cognitions and/or functions are presented in Table 1.

*Snodgrass Naming Task (Snodgrass & Vanderwart, 1980)*

Participants name the object of 15 random presented stimuli, obtained from a standardized set of pictures (Snodgrass & Vanderwart, 1980). The test instructions and scoring methods that are used are as following: Responses are incorrect when image is incorrectly named or response time is longer than three seconds, considering that healthy participants must be able to give a correct answer in that timeframe (Kaplan, Goodglass & Weintraub, 2001). For each stimulus, the correct answers are predetermined (see: Appendix I). The sum of the total score for each participant is calculated (1=correct, 0=incorrect). Subsequently, the mean score and standard deviation per version in the total group was calculated.

*Reading*

Participants are instructed to read out loud five written words, five sentences and two short stories in common reading pace. The raw score, which are defined as the incorrect pronounced words, will be noted per participant. The total number of incorrect pronounced words per condition is calculated.

*Stroop Test with and without block (Delis, Kaplan & Kramer, 2001)*

The Stroop Test assesses the ability to inhibit cognitive interference that occurs when the processing of a specific stimulus impedes the simultaneous processing of a second stimulus. Participants must inhibit a dominant response and name the color in which a word is printed, while ignoring the written word itself. When the stimulus is presented with a block around the written word, the written word must be read. A shortened form in which three times 12 stimuli is being presented in both conditions, without breaks. The total time of the 3 x 12 blocks is

noted in seconds, separately for the condition with and without blocks. The number of errors and corrections is noted. The mean of total seconds, errors and corrections per version is calculated separately for both conditions.

*Digit Span Forward and Backward (Wechsler, 2008)*

The Digit Span consist of a Forward condition and a Backward condition and aims to be a (working) memory test. Participants must repeat three sequence of numbers, while the sequence increases in length (3-4-5). In the Digit Span Backward condition, participants had to repeat a sequence of numbers in the reverse order. This sequence also increases in length (2-3-4). Per participant the raw score is noted, which is indicated as the highest achieved span in the forward and backward condition separately. Subsequently, the mean span and standard deviation for both conditions are calculated in the four versions.

*Dot Counting Test with background (James, Plant & Warrington, 2001)*

The Dot Counting Test with background is a test for visual perception and offers the possibility to detect an occurring neglect. Participants are instructed to count the dots as quickly as possible on five given stimuli, varying from three to five dots. The added background to the standard test, on which the dots are presented is added to the standard test and provides extra distraction. Incorrect when wrong number is said or when the response time is longer than two seconds given that healthy participants should be able to give a correct answer in that timeframe (Herrera-Guzmán, Peña-Casanova, Lara, Gudayol-Ferre & Böhm, 2004). Per participant, the raw score is noted, which is defined as the total errors. The mean error per stimuli in the total group is calculated per version.

*How are They Feeling?: Colorcards (Speechmark Publishing Ltd, 2012)*

Photographs of three emotionally expressive situations are presented and participants are asked to name the emotions of the people in the picture. Participants are explicitly asked not to give a situation description of the presented stimuli, but to name one specific emotion which the designated person in the picture can experience. Given that there are many variations in answers, the correct emotions per photograph is determined (see: Appendix 2 ). The raw score per participant is calculated (1=correct; 2=incorrect answer). The sum of the correct answers in the total group for the 12 stimuli (three in each version) is calculated.

**Table 1.** Neuropsychological tests and corresponding cognitions and/or functions

| ***Cognitive Functions*** | ***Neuropsychological Tests*** |
|---|---|
| *Language* | *Snodgrass Naming Task, Reading* |
| *Memory* | *Digit Span Forward and Backward* |
| *Attention, Executive Functioning* | *Stroop Test with and without Block,* |
| *Visual perception* | *Dot Counting Test with Background* |
| *Emotion recognition* | *How are They Feeling?: Colorcards* |

### *Statistical Analyses*

For the statistical analyses, IBM SPSS Statistics 25 was used. Raw scores were used in all analyses. To explore whether mean performances of the four cognitive screeners differ in each neuropsychological task, a series one-way analysis of variance are performed. If the assumption of normality in the data could be assumed, mean performances were compared with one-way Analysis of Variance (ANOVA). In case of significant group differences, Tukey post-hoc test will be performed to indicate were the significant difference exist (Abdi & Williams, 2010). Non-parametric tests will be performed when the assumption of normality was rejected. Kruskal-Wallis one way Analysis of Variance will be used to determine if statistically significant differences exist between the four groups on neuropsychological performance in each domain. If the overall omnibus H-test was significant, multiple pairwise comparisons were conducted to locate the source of the significant difference between group. Non-parametric post-hoc tests will be performed with a Mann-Whitney U-test (Mann & Whitney, 1947). This post-hoc test is essentially a non-parametric two samples t-test and allows for the examination of how the groups differ.

## Results

### *Demographic and clinical characteristics*

All participants (*n* = 38*)* completed the four cognitive screeners. Demographics and clinical characteristics are presented in Table 2. Mean age was 38.34 (SD = 17.63). There was a small female predominance (52.6%) in the sample. Educational level (Verhage, 1983) was normally distributed with a predominance of a middle educational level (M = 5.0; SD = 0.9).

**Table 2.** Demographic and clinical characteristics of the sample (*n* = 38)

| ***Characteristics*** | |
|---|---|
| Sex, *n* (%) | |
| Male | 18 (47.4%) |
| Female | 20 (52.6%) |
| Age in years M ± SD | 38.34 ± 17.63) |
| Range | 18-74 |
| Education level [a], *n* (%) | |
| Low | 12 (34.2%) |
| Middle | 14 (68.4%) |
| High | 12 (31.5%) |
| M ± SD; range | 5.0 ± 0.9; 3-7 |
| Median (IQR) | 5.0 (4.0-6.0) |
| Attention Deficit Hyperactivity Disorder, *n* (%) | 1 (2.6%) |
| Dyslexia, *n* (%) | 1 (2.6%) |
| Colorblindness, *n* (%) | 2 (5.3%) |

[a] Education was classified according to the coding system of Verhage (Verhage, 1983) ranging from 1 (less than six years of primary education) to 7 (University degree)

***Neuropsychological tests***

1. *Snodgrass Naming Task*

Statistical analysis employed non-parametric tests, because of the violation of the normality assumption. Kruskal-Wallis one way ANOVA of the four groups showed a significance difference between groups (H(3) = 10.218, $p$ = .017). Post-hoc Mann-Whitney U test showed that there was a significant difference (U = 551, $p$ = .004) between the total mean score of version 1 compared to version 3. No significant differences in mean score between the other versions are found.

2. *Stroop Test without block and Stroop Test with block*

*2.1 Stroop Test without block*

Reaction time in seconds were right-skewed, therefore natural log (Ln) transformations were performed. Consequently, ANOVA was conducted to compare the mean time in seconds between version 1, version 2, version 3 and version 4. There was a statistically significant differences between groups in reaction time in seconds as demonstrated by one-way ANOVA (F(3,148) = 3.191, $p$ = .025). Tukey post hoc test showed that version 3 was significantly higher than version 4 ($p$ = .039). There was no significant difference between version 1 and 2 ($p$ = .970), version 1 and 3 ($p$ = .050), version 1 and 4 ($p$ = 1.000), version 2 and 3 ($p$ = .141), version 2 and 4 ($p$ = .950). To conduct whether the four versions, differ in the total amount of errors and correction, Kruskal-Wallis one way ANOVA was performed to investigate differences in corrections and errors between the four tests. No significant differences were found between the four versions in errors (H(3) = .970, $p$ = .808) and corrections (H(3) = .656, $p$ = .884). Tukey post hoc test results are presented in Table 3.

**Table 3.** Tukey post hoc test results for Stroop Test without Block (*n* = 4)

| Version | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | - | - | - | - |
| 2 | .970 | - | - | - |
| 3 | .050 | .141 | - | - |
| 4 | 1.000 | .950 | .039* | - |

**: Difference between the two groups was statistically significant at (p = .005)*

*2.2 Stroop Test with block*

Reaction time in seconds were right-skewed, therefore natural log (Ln) transformations were performed. A one-way ANOVA was conducted to compare the mean time in seconds between version 1, version 2, version 3 and version 4. No statistically significant difference between groups is found (F(3,148) = .052, *p* = .984). Kruskal-Wallis one way ANOVA of the four tests was conducted to investigate significant differences errors and corrections between the four versions. No significant differences when comparing the total amount of errors (H(3) = .093, *p* = .993) and corrections (H(3) = 2.079, *p* = .556) between groups.

*3. Digit Span Forward and Backward*

*3.1 Digit Span Forward*

Statistical analysis employed non-parametric tests, because of the violation of the normality assumption. Kruskal-Wallis one way ANOVA of the four groups showed no significance difference in mean span between groups (H(3) = 2.202, *p* = .531).

*3.2 Digit Span Backward*

Statistical analysis employed non-parametric tests, because of the violation of the normality assumption. Kruskal-Wallis one way ANOVA of the four groups showed no significance difference in mean span between groups (H(3) = 6.775, *p* = .079).

## 4. *Reading*

### *4.1 Words*

No statistical analyses are performed in this condition, given that no subjects made errors in each of the four conditions.

### *4.2 Sentences*

No statistical analyses are performed in this condition, given that no subjects made errors in each of the four conditions.

### *4.3 Stories*

Statistical analysis employed non-parametric tests, because of the violation of the normality assumption. Kruskal-Wallis one way ANOVA of the 8 stories showed a significance difference between mean errors (H(7) = 20.536, $p$ = .005).

Post-hoc Mann-Whitney U test indicated that mean error was significantly greater for story 6 than for story 1 (U = 589 , $p$ = .014), story 2 (U = 608, $p$ = .043), story 3 (U = 570, $p$ = .003), story 4 (U = 608, $p$ = .043), story 5 (U = 570, $p$ = .003), story 7 (U = 608, $p$ = .043) and story 8 (U = 608, $p$ = .043). No other significant differences between stories where found. Post hoc Mann-Whitney U test results are presented in Table 4.

**Table 4.** Post hoc Mann-Whitney U test results for Stories ($n$ = 8)

| Story | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | - | - | - | - | - | - | - | - |
| 2 | .558 | - | - | - | - | - | - | - |
| 3 | .317 | .155 | - | - | - | - | - | - |
| 4 | .558 | 1.000 | .155 | - | - | - | - | - |
| 5 | .079 | .155 | 1.000 | .155 | - | - | - | - |
| 6 | .014* | .043* | .003* | .043* | .003* | - | - | - |
| 7 | .307 | .646 | .079 | .646 | .079 | .043* | - | - |
| 8 | .307 | .646 | .079 | .646 | .079 | .043* | - | - |

**: Difference between the two groups was statistically significant at (p = .005)*

5. *Dot Counting Test with background*

No statistical analyses are performed, considering no subject gave an incorrect response and/or the response time was longer than two seconds on the five stimuli in each condition.

6. *How are They Feeling?: Colorcards*

Statistical analysis employed non-parametric tests, because of the violation of the normality assumption. Kruskal-Wallis one way ANOVA for the showed a significant difference between the 12 cards (H(11) = 77.291 $p$ = .000). Post hoc -Mann Whitney test indicated that the answers between multiple stories significantly differ. In Table 5, the results of the Wilcoxon-Mann-Whitney signed rank test are demonstrated.

**Table 5.** Post hoc Wilcoxon-Mann-Whitney signed rank test results for How are They Feeling?: Colorcards ($n$ = 12)

| Card | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | - | - | - | - | - | - | - | - | - | - | - | - |
| 2 | .079 | - | - | - | - | - | - | - | - | - | - | - |
| 3 | .000* | .000* | - | - | - | - | - | - | - | - | - | - |
| 4 | .001* | .034* | .096 | - | - | - | - | - | - | - | - | - |
| 5 | .006* | .178 | .014* | .412 | - | - | - | - | - | - | - | - |
| 6 | 1.000 | .079 | .000* | .001* | .006* | - | - | - | - | - | - | - |
| 7 | .006* | .178 | .014* | .412 | 1.000 | .006* | - | - | - | - | - | - |
| 8 | .317 | .307 | .000* | .004* | .026* | .317 | .026* | - | - | - | - | - |
| 9 | 1.000 | .083 | .000* | .001* | .006* | 1.000 | .006* | .324 | - | - | - | - |
| 10 | .000* | .000* | 1.000 | .096 | .014* | .000* | .014* | .000* | .000* | - | - | - |
| 11 | .003* | .105 | .029* | .592 | .775 | .003* | .775 | .014* | .003* | .029* | - | - |
| 12 | .022* | .458 | .003* | .152 | .532 | .022* | .532 | .091 | .023* | .003* | .364 | - |

**: Difference between the two groups was statistically significant at (p = .005)*

An overview of the discussed results of the between-group comparison for neuropsychological tests are described in Table 6.

**Table 6.** Between-group comparison for each neuropsychological test

| | **H** | ***df*** | ***F-value*** | ***p-Value**** |
|---|---|---|---|---|
| Snodgrass Naming Task [a] | 10.218 | 3 | | .017* |
| Stroop [b] | | | | |
| Stroop without block [b] | - | 3 | 3.191 | .025* |
| Errors [a] | .970 | 3 | | .808 |
| Corrections [a] | .656 | 3 | | .884 |
| Stroop with block [b] | | | .052 | .984 |
| Errors [a] | .093 | 3 | | .993 |
| Corrections [a] | 2.079 | 3 | | .556 |
| Digit Span [a] | | | | |
| Digit Span Forward | 2.202 | 3 | | .531 |
| Digit Span Backward | 6.775 | 3 | | .079 |
| Reading [a] | | | | |
| Words | - | - | | - |
| Sentences | - | - | | - |
| Stories | 20.536 | 7 | | .005* |
| Dot counting with background [a] | - | - | | - |
| How are They Feeling: Colorcards [a] | 77.291 | 11 | | .000* |

**: Difference between the four groups was statistically significant at (p = .005)*

[a] *Kruskal-Wallis one way ANOVA test*

[b] *One-way Analysis of Variance (ANOVA)*

**H**: test statistics for Kruskal-Wallis test, ***df***: degree of freedom

## Discussion

This study examined whether it is possible to expand the current cognitive screener that has been in use over 10 years at the UMC Utrecht Department of Neurosurgery for pre- and intra-operative to allow a broader cognitive screening during awake brain tumor surgery. Part of this objective was conducting reliable parallel versions to monitor cognition over time without the occurrence of possible practice effects (Benedict & Zgaljardic, 1998; Duff et al., 2007).

Analyses were performed separately for each cognitive domain and its corresponding tests. Results shows that it is possible to extent the current cognitive screener in a feasible way in five cognitive domains and functions: language, memory, attention, executive functioning and visual perception. Unfortunately, significant differences have been found between the How Are They Feeling?: Colorcards and therefore cannot be implemented in the tests. Results indicated that three reliable versions can be implemented without any modifications for use during awake brain tumor surgery and with exclusion of the How Are They Feeling?: Colorcards,. Besides, these cognitive screeners could either be administered before and after surgery, thereby minimizing the chance of a possible practice effect due to various stimuli in each version (Benedict & Zgaljardic, 1998; Duff et al., 2007).

Several shortcomings at test level have been observed. First, the group differences between two versions of the Snodgrass Naming Task, a task that aims to objectifies object naming in total score per group, can be explained by a common mistake that has been made in the first version among multiple subjects. In this version, the image of a ‘rugby ball’ was presented and analysis revealed a high margin of error among subjects whereas they gave the wrong (‘football’, ‘basketball’) or unspecific (‘ball’) answer. No differences between groups have been found when exclusion this image. Replacement of this specific picture for a more unambiguous picture should be considered and could eliminate the significant difference between versions while maintaining the same number of stimuli in each version. However, new data collection and analysis must determine whether this new picture is reliable.

Secondly, analysis of variance of the Stroop Task without Block, a task that aims to measure attention and executive functioning, showed a significant difference in mean reaction time in seconds between two versions. This group differences can be explained by the fact that the majority of the subjects started with a specific version, despite the versions where administered in random order. An explanation for this finding could be the occurrence of *test wiseness* (TW). This phenomenon is defined as a test-staking strategy that can be taught by instruction and that the knowledge that has been gained will enable a subject to obtain a higher

score than at previous stimuli in repeated measures (Wahlstrom & Boersma, 1968).

Despite the possible appearance of *test wiseness* (TW), the Stroop Task that has been used is assumed to be suitable for repeated testing and provide reliable results with concerning executive functioning and attention (Jensen, 1965). Furthermore, only the color-word interference condition of the Stroop Task is presented among the subjects since intra-operative test assessments must be short in presentation time (Coello et al., 2013). Originally, the complete Stroop task consist of three parts and aims to measure the '*stroop-effect*', which is defined as the delay between congruent (color and word correspondent) and incongruent stimuli (color and word do not correspondent) and refers to the mismatch between the name of the presented color and the color it is printed on (Jensen, 1965; Delis, Kaplan & Kramer, 2001). The Stroop Task that has been used in the current cognitive screener only assess the ability to inhibit cognitive interference that occurs when processing of a specific stimulus feature impedes the simultaneous processing of a second stimulus attribute. Therefore, the cognitive interference cannot be calculated and serves as an indication of executive functioning and attention compared to pre-operative functioning for noticing possible changes in cognition.

Nevertheless, literature states that when there is a high probability of occurrence of executive functioning problems, for example in frontal gliomas, an extensive Stroop Task is available to assess during operations to obtain an objective indication of executive functioning and attention during awake brain tumor surgery (Puglisi et al., 2018). This intra-operative Stroop Task consist of more stimuli in the color-word interference condition and offers the possibility to compare patients' score with norm scores obtained from healthy controls. Although the assessment of executive functioning seems difficult, the current Stroop Task that has been used in the test provides a reliable and sensitive instrument for use during operations (Wager et al., 2013).

The tasks concerning reading words and sentences yields reliable results among subjects. No analyses were performed due to achieving errorless raw scores in each version in all participants. This result was also achieved in the Dot Counting Test with background (James, Plant & Warrington, 2001). Subjects in this study have scores that are the upper limit of the tests, this is the so called c*eiling effect* (Everitt, 2002). The critical comment in this case is whether the tests are sensitive enough to objectify subtle changes in cognition. However, literature states that healthy participants should be able to perform on this task without errors, given that the Dot Counting Test has been used to assess difficulties with spatial scanning or localization of a single point in space and therefore a possible neglect could be observed (Adlington, Laws & Gale, 2009; Economou & Papageorgiou, 2011). Hence, it was expected

that healthy participants could complete the tasks without errors and indicates reliable and valid data. As mentioned before, these versions have now been tested in healthy controls. However, these tests will be used for patients with cognitive impairment whose severity and complaints can diverge. For that reason, it is important to have different levels of difficulty in order to fall back on easier stimuli if necessary, for instance reading out sentences instead of short stories. However, the subtasks stories showed significant group differences between two stories. This difference can be explained by the word '*gediagnosticeerd*' (diagnosed) that was difficult for a large part of the subject. Removing this specific word in the story, no significant differences has been found. However, new analysis and addition of another word should indicate whether this adjustment is reliable.

Furthermore, this approach of assessing language during awake brain surgery is still restricted. Language is a more extensive domain and this cognitive screener does not cover the integral variety of expressive and receptive language functions that are necessary for adequate communication (De Witte et al., 2015). Tests for the assessment of language functions during awake brain surgery are widely reported (Ruis, 2018). Besides, traditionally awake brain tumor surgery is done in the language area in the dominant hemisphere (Coello et al., 2013) and therefore much attention is paid to intra-operative cognitive assessments with focus on language area (Grossman & Ram, 2013; Ruis, 2018). Possibilities to conduct patient-oriented test batteries concerning the language domain are already in use when required and can be added to the current cognitive screener, for instance the standardized Dutch Linguistic Test Battery (De Witte et al., 2015).

In contrast, tests for monitoring emotions during awake brain surgery are reported in a minority of studies and are still lacking (Ruis, 2018). Therefore, we implemented How are They Feeling?: Colorcards, where subjects were asked to name the expressed emotions in pictures. However, this task is not reliable to implement in the cognitive screener due to significant differences between cards. This result can be explained by the fact that faces of the described person were not visible on all cards while literature shows that emotions are best recognized by facial expressions in comparison with body gestures and language (Bänziger, Granjean & Scherer, 2009). This requires participants to use the environment or situation where the person on the picture is presented to label the emotion. As a result, stimuli were therefore ambiguous and a high variability in answers occurred. Other emotion recognition tasks, for instance the Emotion Recognition Task (Giussani, Pirillo & Roux, 2010) and Reading the Mind in the Eyes Test (Baron-Cohen, Wheelwright, Hill, Raste & Plumb, 2001) appears to provide reliable results for implementing during awake brain tumor surgery. Due to little research, further

investigations must indicate whether this can be implemented on large scale (Coello et al., 2013).

Lastly, due to the small sample size in the current study, the assumption of normality could not be assumed, and analysis showed that the assumption for normality was rejected in multiple domains. Non-parametric Kruskal-Wallis tests has to be used, which is less powerful than a One-Way Analysis of Variance (Hecke, 2012). Results should therefore be interpreted with caution.

The current study has the following strengths. First, the recruitment of participants in personal environment of the researcher leads to a high availability of subjects. Secondly, a high variability in participants concerning educational level, age and gender has been conducted, which lead to normally distributed factors and reflects the population as a whole. A few participants mentioned to be diagnosed with dyslexia or color-blindness, but no effects due to these factors have been found. Furthermore, the constructed parallel versions are suitable for implementing pre, intra-and post-operative for assessing cognitions in long-term surviving patients with brain tumors. This is crucial, because that multimodal cognitive decline is being recognized as an independent prognostic factor in patients with brain tumors and appears to be a major indicator for tumor regrowth after treatment (Lieberman et al., 1982; Taphoorn & Klein, 2004).

Considered that individual patient care is required, and standardized neuropsychological assessment cannot be administrated during surgery, these cognitive screeners can help clinicians to gain knowledge in general cognitive functioning in multiple timeframes (Coello, et al., 2013). Moreover, pre-operative assessment of the current cognitive screener serves as constructing a baseline indication of patients' cognitions and makes it possible to compare pre-operative and intra-operative functioning. Therefore, it should be kept in mind that the assessment of patients' score is mainly subjective, considering that intra-operative observations according to tasks performance are of crucial importance to achieve an optimal onco-functional balance in patients (Coello et al., 2013; Duffau & Mandonnet, 2013). However, more extensive reliable test material makes it feasible to compare patients' results with standard norm scores and thus provides more information about cognitive functioning in brain tumor patients compared to healthy persons. Determination in which domains extensive test material is desired can be done by information obtained from other studies concerning the relationship between tumor location and cognitions. This way, optimal cognitive monitoring before, during and after awake brain tumor surgery can be achieved and can contribute to further research in the relationship between tumor location and cognitions (Coello et al., 2013).

In conclusion, the expanded cognitive screeners are a reliable outcome measure to be implemented during awake brain tumor surgery to detect subtle changes in a broad spectrum of cognitions. Furthermore, the constructed reliable parallel versions make it possible to assess cognitive functioning over time and provides long term information about the course of cognitive functioning with minimizing the possible occurrence of a practice effect. Besides, a task for emotion recognition which could be implemented during awake brain tumor surgery is lacking and desired, since deficits in emotion recognition is a common cognitive consequence of certain brain tumors (Andrewes et al., 2003).

## References

Abdi, H., & Williams, L. J. (2010). Tukey's honestly significant difference (HSD) test. *Encyclopedia of Research Design.* Thousand Oaks, CA*: Sage*, 1-5.

Adlington, R. L., Laws, K. R., & Gale, T. M. (2009). Visual processing in Alzheimer's disease: Surface detail and colour fail to aid object identification. *Neuropsychologia*, *47*(12), 2574-2583. doi:10.1016/j.neuropsychologia.2009.05.004

Andrewes, D. G., Kaye, A., Murphy, M., Harris, B., Aitken, S., Parr, C., & Bates, L. (2003). Emotional and social dysfunction in patients following surgical treatment for brain tumour. *Journal of Clinical Neuroscience*, *10*(4), 428-433. doi:10.1016/S0967-5868(03)00086-9

Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: the Multimodal Emotion Recognition Test (MERT). *Emotion*, *9*(5), 691. doi:10.1037/a0017088

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241-251. doi:10.1111/1469-7610.00715

Benedict R.H.B. & Zgaljardic D.J. (1998). Practice effects during repeated administrations of memory tests with and without alternative forms. *Journal of clinical and experimental neuropsychology 20*: 339-352.

Cockrell, J. R., & Folstein, M. F. (2002). Mini-mental state examination. *Principles and practice of geriatric psychiatry*, 140-141. doi:10.1002/0470846410.ch27(ii)

Coello, A. F., Moritz-Gasser, S., Martino, J., Martinoni, M., Matsuda, R., & Duffau, H. (2013). Selection of intra-operative tasks for awake mapping based on relationships between tumor location and functional networks: A review. *Journal of neurosurgery*, *119*(6), 1380-1394.

Conte, V., Baratta, P., Tomaselli, P., Songa, V., Magni, L., & Stocchetti, N. (2008). Awake neurosurgery: an update. *Minerva anestesiologica*, *74*(6), 289-292.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive function system: examiners manual*. Psychological Corporation.

De Witte, E., Satoer, D., Robert, E., Colle, H., Verheyen, S., Visch-Brink, E., & Mariën, P. (2015). The Dutch Linguistic Intra-operative Protocol: A valid linguistic approach to awake brain surgery. *Brain and Language*, *140*, 35-48.

Duffau, H. (2010). Awake surgery for nonlanguage mapping. *Neurosurgery*, *66*(3), 523-529. doi:10.1227/01.NEU.0000364996.97762.73

Duffau, H., & Mandonnet, E. (2013). The “onco-functional balance” in surgery for diffuse low grade glioma: integrating the extent of resection with quality of life. *Acta neurochirurgica*, *155*(6), 951-957. doi:10.1007/s00701-013-1653-9

Dziedzic, T., & Bernstein, M. (2014). Awake craniotomy for brain tumor: indications, technique and benefits. *Expert review of neurotherapeutics*, *14*(12), 1405-1415. doi:10.1586/14737175.2014.979793

Economou, A., & Papageorgiou, S. (2011). Discrepancies of memory functions in healthy and memory-impaired older persons. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *7*(4), S275-S276. doi:10.1016/j.jalz.2011.05.789

Everitt, B. S. (2002). *The Cambridge dictionary of statistics* (2nd ed.). Cambridge, UK: CUP. doi:10.1017/CBO9780511779633

Freyschlag, C. F., & Duffau, H. (2014). Awake brain mapping of cortex and subcortical pathways in brain tumor surgery. *Journal of neurosurgical sciences*, *58*(4), 199-213.

Grossman, R., & Ram, Z. (2013). awake Craniotomy in Glioma surgery. *European Association of NeuroOncology Magazine*, 4(1), 27–33.

Giussani, C., Pirillo, D., & Roux, F. E. (2010). Mirror of the soul: a cortical stimulation study on recognition of facial emotions. *Journal of neurosurgery*, *112*(3), 520-527. doi:10.3171/2009.5.JNS081522

Heberlein, A. S., Padon, A. A., Gillihan, S. J., Farah, M. J., & Fellows, L. K. (2008). Ventromedial frontal lobe plays a critical role in facial emotion recognition. *Journal of cognitive neuroscience*, *20*(4), 721-733. doi:10.1162/jocn.2008.20049

Hecke, T. V. (2012). Power study of anova versus Kruskal-Wallis test. *Journal of Statistics and Management Systems*, *15*(2-3), 241-247. doi:10.1080/09720510.2012.10701623

Herrera-Guzmán, I., Peña-Casanova, J., Lara, J. P., Gudayol-Ferre, E., & Böhm, P. (2004). Influence of age, sex, and education on the Visual Object and Space Perception Battery (VOSP) in a healthy normal elderly population. *The Clinical Neuropsychologist*, *18*(3), 385-394. doi:10.1080/1385404049052421

de Jager, C. A., Hogervorst, E., Combrinck, M., & Budge, M. M. (2003). Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. *Psychological medicine*, *33*(6), 1039 1050. doi:10.1017/S0033291703008031

James, M., Plant, G. T., & Warrington, E. K. (2001). *CORVIST: Cortical vision screening test: manual & test materials*. Thames Valley Test Company.

Jensen, A. R. (1965). Scoring the Stroop test. *Acta psychologica*, *24*(5), 398-408. doi:10.1016/0001-6918(65)90024-7

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston naming test*. Pro-ed.

Lieberman, A. N., Foo, S. H., Ransohoff, J., Wise, A., George, A., Gordon, W., & Walker, R. (1982). Long term survival among patients with malignant brain tumors. *Neurosurgery*, *10*(4), 450-453. doi:10.1227/00006123-198204000-00006

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 50-60.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695-699. doi:10.1111/j.1532-5415.2005.53221.x

Puglisi, G., Sciortino, T., Rossi, M., Leonetti, A., Fornia, L., Nibali, M. C., ... & Bello, L. (2018). Preserving executive functions in nondominant frontal lobe glioma surgery: an intra-operative tool. *Journal of neurosurgery*, *131*(2), 474-480.

Ruis, C. (2018). Monitoring cognition during awake brain surgery in adults: A systematic review. *Journal of clinical and experimental neuropsychology*, *40*(10), 1081-1104 doi:10.1080/13803395.2018.1469602

Sala, F., & Lanteri, P. (2003). Brain surgery in motor areas: the invaluable assistance of intra operative neurophysiological monitoring. *Journal of neurosurgical sciences*, *47*(2), 79.

Sanai, N., Mirzadeh, Z., & Berger, M. S. (2008). Functional outcome after language mapping for glioma resection. *New England Journal of Medicine*, *358*(1), 18-27. doi:10.1056/NEJMoa067819

Santini, B., Talacchi, A., Squintani, G., Casagrande, F., Capasso, R., & Miceli, G. (2012). Cognitive outcome after awake surgery for tumors in language areas. *Journal of neuro-oncology*, *108*(2), 319-326. doi:10.1007/s11060-012-0817-4

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory*, *6*(2), 174. doi:10.1037/0278-7393.6.2.174

Speechmark. (2012). *How are They Feeling?: Colorcards.* United Kingdom: Taylor & Francis Ltd.

Taphoorn, M. J., & Klein, M. (2004). Cognitive deficits in adult patients with brain tumours. *The Lancet Neurology*, *3*(3), 159-168. doi:10.1016/S1474-4422(04)00680-5

Van Zandvoort, M., Ruis, C., & Hendriks, M. (2016). Wakkere hersenoperaties: de klinisch neuropsychologisch aspecten. *Neuropraxis*, *20*(3), 91-95. doi:10.1007/s12474-016-0126-3

Verhage, F. (1983). Het coderen van het opleidingsniveau voor researchdoeleinden [Educational classification system for research purposes: revised version]. *Internal publication Groningen, the Netherlands: Academic Hospital Groningen, State University Groningen*.

Wager, M., Du Boisgueheneuc, F., Pluchon, C., Bouyer, C., Stal, V., Bataille, B., ... & Gil, R. (2013). Intra-operative monitoring of an aspect of executive functions: administration of the Stroop test in 9 adult patients during awake surgery for resection of frontal glioma. *Operative Neurosurgery*, *72*(2), ons169-ons181.

Wahlstrom, M., & Boersma, F. J. (1968). The influence of test-wiseness upon achievement. *Educational and Psychological Measurement*, *28*(2), 413-420. doi:10.1177/001316446802800221

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–fourth edition (WAIS–IV)*. San Antonio, TX: NCS Pearson.

Whittle, I. R., Midgley, S., Georges, H., Pringle, A. M., & Taylor, R. (2005). Patient perceptions of “awake” brain tumour surgery. *Acta Neurochirurgica*, *147*(3), 275-277. doi:10.1007/s00701-004-0445-7

Yoshii, Y., Tominaga, D., Sugimoto, K., Tsuchida, Y., Hyodo, A., Yonaha, H., & Kushi, S. (2008). Cognitive function of patients with brain tumor in pre-and postoperative stage. *Surgical neurology*, *69*(1), 51-61. doi:10.1016/j.surneu.2007.07.064

# Appendix

## *I: Answer Sheet Snodgrass Naming Task*

*An overview of the correct responses in the Snodgrass Naming Task per version*

| | *Version 1* | *Version 2* | *Version 3* | *Version 4* |
|---|---|---|---|---|
| *Stimuli* | | | | |
| *1.* | Taart | Gieter | Vlieger | Jas, mantel |
| *2.* | Kameel, dromedaris | Boom | Sleutel | Ballon |
| *3.* | Kaars | Zebra | Wiel, rad | Ton, barrel |
| *4.* | Ketting | Horloge | Paard | Vlinder |
| *5.* | Stoel | Schildpad | Sneeuwpop, sneeuwman | Anker |
| *6.* | Kers | Tandenborstel | Olifant | Bank |
| *7.* | Kanon, mortier | Zwaan | Vlag, rode vlag | Slak |
| *8.* | Auto | Bril | Zon | Gitaar |
| *9.* | Kerk | Sok | Schommel | Citroen |
| *10.* | Envelop, brief | Tomaat | Kip | Motor |
| *11.* | Hek | Kruk | Vliegtuig | Liniaal |
| *12.* | Vis | Verkeerslicht, stoplicht | Peer | Riem |
| *13.* | Clown | Maan, halve maan | Strik | Kikker |
| *14.* | Mokje, kopje | Paraplu | Molen, windmolen | Bloem, viool |
| *15.* | Rugbyball, rugby | Tas, schoudertas | Wasknijper, knijper | Ster |

## *II: Answer Sheet Stroop Task without Block*

*Version 1*

| | | | | | |
|---|---|---|---|---|---|
| Rood | Rood | Blauw | Geel | Rood | Rood |
| Groen | Rood | Geel | Rood | Blauw | Blauw |
| Groen | Blauw | Blauw | Geel | Blauw | Geel |
| Rood | Geel | Geel | Groen | Groen | Rood |
| Geel | Blauw | Geel | Blauw | Rood | Rood |
| Rood | Geel | Rood | Groen | Rood | Rood |

*Version 2*

| | | | | | |
|---|---|---|---|---|---|
| Geel | Rood | Rood | Rood | Blauw | Blauw |
| Groen | Geel | Groen | Groen | Blauw | Geel |
| Groen | Rood | Blauw | Geel | Groen | Groen |
| Blauw | Blauw | Rood | Rood | Groen | Blauw |
| Groen | Blauw | Groen | Blauw | Geel | Geel |
| Geel | Geel | Blauw | Geel | Blauw | Rood |

*Version 3*

| | | | | | |
|---|---|---|---|---|---|
| Blauw | Geel | Groen | Blauw | Blauw | Blauw |
| Groen | Geel | Geel | Rood | Groen | Blauw |
| Geel | Rood | Blauw | Groen | Rood | Blauw |
| Geel | Groen | Geel | Blauw | Rood | Blauw |
| Geel | Blauw | Geel | Blauw | Rood | Blauw |
| Rood | Geel | Rood | Groen | Rood | Rood |

*Version 4*

| | | | | | |
|---|---|---|---|---|---|
| Geel | Blauw | Geel | Blauw | Rood | Rood |
| Rood | Geel | Rood | Groen | Rood | Rood |
| Geel | Rood | Rood | Rood | Blauw | Blauw |
| Groen | Geel | Groen | Groen | Blauw | Geel |
| Groen | Blauw | Groen | Blauw | Geel | Geel |
| Geel | Geel | Blauw | Geel | Blauw | Rood |

## *III: Answer Sheet Stroop Task with Block*[1]

*Version 1*

| | | | | | |
|---|---|---|---|---|---|
| **Rood** | Rood | Blauw | Geel | Rood | Rood |
| Geel | Groen | Geel | Blauw | Rood | **Rood** |
| Groen | **Rood** | Blauw | Geel | Blauw | Geel |
| Rood | Geel | **Blauw** | Groen | Groen | Rood |
| Rood | Rood | Blauw | Geel | Rood | Rood |
| Groen | Rood | **Groen** | **Geel** | Blauw | Blauw |

*Version 2*

| | | | | | |
|---|---|---|---|---|---|
| Groen | Blauw | Blauw | Geel | Blauw | Geel |
| Rood | **Blauw** | Geel | **Geel** | Groen | Rood |
| Geel | **Rood** | Geel | Blauw | Rood | Rood |
| Rood | Geel | Rood | Groen | Rood | **Geel** |
| Blauw | Geel | Groen | Blauw | **Groen** | Blauw |
| Groen | Geel | **Rood** | Rood | Groen | Blauw |

*Version 3*

| | | | | | |
|---|---|---|---|---|---|
| **Rood** | Rood | Rood | **Groen** | Blauw | Blauw |
| Groen | Geel | Groen | Groen | Blauw | Geel |
| Groen | **Rood** | Groen | Blauw | Geel | Geel |
| **Rood** | Geel | Blauw | Geel | Blauw | Rood |
| Groen | Rood | **Groen** | Geel | **Geel** | Groen |
| Blauw | Blauw | Rood | Rood | Groen | Blauw |

*Version 4*

| | | | | | |
|---|---|---|---|---|---|
| **Geel** | Rood | Blauw | Geel | Rood | Rood |
| Groen | Rood | Geel | **Geel** | Blauw | Blauw |
| Geel | Rood | Blauw | Groen | Rood | **Groen** |
| Geel | **Blauw** | Geel | Blauw | Rood | Blauw |
| Blauw | Geel | Groen | Blauw | Blauw | Blauw |
| Groen | **Rood** | Geel | Rood | Groen | **Rood** |

[1] The bold words include a block and has been defined as an incongruent stimulus