Utrecht University

The Behavioural Assessment of the Dysexecutive Syndrome: An Evaluation of the Time Limits for Older Participants and New Dutch Normative Data

Master thesis Neuropsychology

Utrecht University

Name: M. H. Dercksen

Student number: 6506534

Supervisor:

Dr. Carla Ruis

In collaboration with Utrecht University and University Medical Centre Utrecht

Abstract

The BADS is a widely used neuropsychological test battery with a high ecological validity. However, the current Dutch normative data dates back 30 years and several practical issues have arisen. The aims of this study were twofold: performing an extensive evaluation of the time limits of several BADS subtests, and providing new and improved Dutch normative tables for the BADS. It was hypothesized that older participants would need more time to complete the subtests than younger participants, uncovering the need for separate time limits for different age groups. In total, 121 healthy participants were included in this study and divided equally into four groups: Young Adults, Middle-aged Adults, Older Adults, and Elderly. Elderly generally needed more time than the other groups to complete several of the subtests, reaching statistical significance for the Rule Shift Cards, Action Program, and most Zoo Map time variables. However, the majority of Elderly did not violate the current time limits, making Elderly-specific time limits seem unnecessary. Young Adults completed the subtests remarkably faster and without violations. Thus, perhaps the time limits should be stricter for Young Adults, instead of more lenient for Elderly. The new norms were compared with the current norms, demonstrating several significant higher means for the new data. With two clinical cases it was illustrated that the new normative data provides stricter classifications of performances. Overall, this study indicates the need for more extensive research into the current time limits and new Dutch normative data of the BADS.

Introduction

The consequences of brain injury that patients most frequently experience, are difficulties with planning, problem-solving, reasoning, and other aspects of adaptive everyday functioning (Boelen, Spikman, Rietveld, & Fasotti, 2009; Ghawami, Sadeghi, Raghibi, & Rahimi-Movaghar, 2016). The umbrella term for such processes is ‘executive functioning’. Executive functioning concerns a wide range of cognitive processes that are used across a variety of situations (Burgess & Stuss, 2017). As a consequence, executive dysfunction can lead to serious impairments in behaviour, which in turn can be detrimental to someone’s academic performance, job performance, or ability to function in daily life activities (Ghawami et al., 2016). Existing neuropsychological tests of executive functioning, such as the Wisconsin Card Sorting Test and the Stroop Test, did not possess an adequate level of ecological validity – which is the measure of the combination of complex skills used in real-life tasks. For this reason, a neuropsychological test battery was developed in 1996 to resolve this issue: the Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996).

The BADS consists of two questionnaires and six subtests: Rule Shift Cards, Action Program, Key Search, Temporal Judgment, Zoo Map, and Modified Six Elements. Across these subtests, participants need to organise, plan, and adjust behaviour and make estimations in order to answer questions or solve problems. For each subtest the raw score is converted into a profile score ranging from 0-4. An overall profile score is computed by summing the individual profile scores of the six subtests. The overall profile score can be converted into a standardized score, with a mean of 100 and a standard deviation of 15, which allows for comparison with scores of other tests. The original normative data of the BADS was compiled with the use of English norm groups, but for the Dutch version of the BADS new normative data has been collected among Dutch participants (Krabbendam & Kalff, 1999). Even though it is demonstrated that the BADS has favourable psychometric properties and a very high ecological validity (Crawford & Henry, 2005; Wilson et al., 1996), some discrepancies regarding the score interpretation have come to light.

The sources of these discrepancies are comparisons with other test scores and observations from clinical neuropsychologists. The discrepancies will be illustrated with the use of two examples from clinical practice (see Box 1 and 2). These examples show that although both patients scored poorly on some subtests of the BADS, their scores could be interpreted as average performances. In addition to these conflicting test scores, several clinicians from the University Medical Center Utrecht (UMCU) have expressed their points of criticism. First of all, the range of the profile scores (0-4) is considered too small, which leads to low differential power – this might be related to the discrepancies described above. Secondly, it is observed that the profile scores are interpreted based on a mean and standard deviation, while it is stated in the supplement of the Dutch version of the BADS that the distribution of performance data was skewed (Krabbendam & Kalff, 1999). The use of several statistics, such as a mean and standard deviation, is most accurate when the variable has a normal distribution. However, when this distribution is skewed, the use of many common statistical techniques will be less valid. This may also be the case for the BADS – especially since the range of the profile score is relatively small. Another point of criticism concerns the Temporal Judgment subtest, during which participants have to estimate the duration of several everyday activities. Currently, this subtest wields a quite strict scoring method, making it impossible to differentiate between answers that just slightly deviate from the correct answer range and answers that deviate considerably. Overall, it is highly important that these issues are to be resolved, given that the BADS is widely used in practice.

There are several possible solutions to the abovementioned issues, such as expanding the range of the profile score, altering the computation method of the normative data (e.g. with the use of percentiles instead of a mean and standard deviation), or converting the raw scores into standardized scores (instead of the profile scores). With regard to the Temporal Judgment subtest, this issue can be solved by increasing the number of points that can be earned, and assigning the points based on how close the guess was to the correct answer range. As a result, two points (instead of one) will be awarded when a given answer falls within the correct answer range. One point (instead of zero) will be awarded when the answer falls just out if this range, but still lies within a particular rang. When the given answer deviates even more from the latter range, zero points will be awarded. Answer ranges can be determined with the use of percentiles. This scoring method will likely enhance the level of specificity of the Temporal Judgment subtest. In the method section will be explained which of these solutions are best suited to the current research.

A 64 year old woman presents herself with slow, progressive cognitive complaints. She made the maximum amount of nine errors on the Rule Shift Cards subtest and earned a profile score of 1. With the current mean of 2.3 and standard deviation of 1.4, this profile score can be converted into a *z*-score of -0.93. This can be interpreted as an average performance. However, she performed poorly on another executive task (Brixton), where her score was at the first percentile, which means she experiences difficulty with mental flexibility.

Box 1. First example from clinical practice.

A 60 year old woman presents herself with word-finding problems and some possible indication for the emergence of dementia. The patient performed poorly on the BADS Zoo Map subtest, where she made many errors in the first condition and broke a rule in the second condition. This resulted in a raw score of 6 and a corresponding profile score of 2. With a mean of 1.9 and a standard deviation of 1.2, this profile score converts to a *z*-score of 0.08, which can be interpreted as an average performance.

Box 2. Second example from clinical practice.

In addition to the suggestions of improvement given by clinicians, this study will focus on another possible point of improvement concerning the speed of performance of older participants. Currently, three subtests of the BADS take into account how much time it takes for someone to complete that task, namely the Rule Shift Cards, Key Search, and Zoo Map. When it takes longer than a predetermined time period to complete a task, one penalty point is deducted from the profile score of that subtest. Currently, the same predetermined time periods hold for all age groups. However, it is well-established in literature and clinical practice that cognitive performance generally declines with age, due to age-associated changes in the brain (Deary et al., 2009; Plumet, Gil, & Gaonac'h, 2005; Salthouse, 2009). In such, Salthouse (2009) demonstrated several negative correlations between age and, among others, speed, spatial visualization, and working memory. Furthermore, the prefrontal cortices seem to be more significantly affected by age-related declines than other neocortical regions (Raz & Rodrigue, 2006). As a result, cognitive functions that are associated with the prefrontal cortex, such as executive functions and speed of processing, start to decline earlier and at faster rates than others, such as verbal ability (Deary et al., 2009; Raz & Rodrigue, 2006; Waters, Sawyer, & Gansler, 2017). Interestingly, Krabbendam and Kalff (1999) also described an age-related decline in performance in the supplement of the Dutch version of de BADS: the oldest age group (64-84 years old) performed substantially worse than the younger age groups. Moreover, this decline in executive performance on the BADS is also illustrated in a more recent study by Burda and colleagues (2017), where performances on the BADS were compared between young, middle-aged, and older participants. This study demonstrated that older participants performed significantly worse than young and middle-aged participants.

Although it is evident that older participants perform more poorly on the BADS than younger participants, it is not explicitly stated that older participants perform more slowly than others – only that this group performs worse than younger participants. Hence, one of the aims of this study is to investigate whether the same time limits should apply to all age groups. It is hypothesized that older participants in this study will not only perform more poorly than the younger participants, but they will also take significantly more time completing the three subtests than the younger groups, and therefore a distinction is needed between the predetermined time limits of the different age groups. Another aim of this study is to collect new normative data among healthy Dutch participants and to compare this new data to the existing Dutch normative data.

Methods

Participants

This study had several inclusion criteria: (1) a minimum age of 18; (2) the absence of neurodegenerative disorders, traumatic brain injury, a history of a (removed) brain tumour or cardiovascular attack, epilepsy, or severe psychopathology (e.g., schizophrenia, personality disorders, etc.); and (3) an adequate proficiency of the Dutch language. Participants were initially recruited through relatives and acquaintances of the researchers, and then further recruited through the social networks of tested participants. All participants voluntarily took part in this study.

Materials

Demographic questionnaire

All participants filled out a demographic questionnaire, consisting of questions about their gender, age, educational level, and whether they were diagnosed with dyslexia, ADHD, or another attention disorder. The latter was included in order to examine whether the presence of such disorders would be of influence on the test performance.

IQ measure

Prior to the administration of the BADS, all participants completed the Dutch version of the Adult Reading Test: the *Nederlandse Leestest voor Volwassenen* (NLV; Schmand, Lindeboom, & van Harskamp, 1992). This test consists of a list of 50 Dutch words and is developed to estimate the level of premorbid intelligence. The total score is converted to an IQ score in accordance with age and gender.

Behavioural Assessment of the Dysexecutive Syndrome

The BADS consists of the following six subtests: Rule Shift Cards, Action Program, Key Search, Temporal Judgment, Zoo Map, and Modified Six Elements. A detailed description of the subtests can be found in Appendix A. The BADS comes with two Dysexecutive Questionnaires (DEX). The DEX consists of 20 items concerning several areas of change (e.g., emotional, personality, motivational, behavioural, cognitive). Usually, both the patient and a close friend or family member fill out the DEX to measure the degree of insight of change or illness. Since the current study only focused on healthy participants, there was no use for the administration of the DEX questionnaires.

Procedure

Potential participants were contacted through phone calls or social media, where they would receive additional information about this study, such as the aim of this study and duration of test administration. Also, it was checked whether these potential participants would meet the inclusion criteria. When participants met these criteria, appointments for the test administration were made.

The majority of the participants were tested at their place of residence. Prior to the test administration, the participants once again received a short explanation of the BADS, the aim of the study and what to expect. Additionally, they were asked to read and sign an informed consent form and to complete the demographic questionnaire. The latter was either done on paper or on an electric device. Thereafter, the NLV was administered. Participants were presented with 50 Dutch words and they were instructed to read the words oud loud at their own pace. Next, the subtests of the BADS were administered in the usual order: Rule Shift Cards, Action Program, Key Search, Temporal Judgment, Zoo Map, and Modified Six Elements. Before each subtest, participants were given the instructions of that subtest and they had the opportunity to ask questions. The duration of a testing session varied from 30 to 60 minutes per participant.

Alterations regarding the scoring methods and statistical analysis

Ideally, the new normative data would be based on percentiles, instead of the current use of the mean and standard deviation of the profile scores. However, the sample size of the current study (ca. 30 participants per age group) does not allow for the use of percentiles, since each age group then would need to contain at least 100 participants. Thus, the decision has been made to base the new normative data on the mean and standard deviation of the raw scores of each subtest. By using the mean and standard deviation of the raw scores instead of the profile scores, the score range generally expands for each subtest (with the exception of the Action Program), which allows for more specific comparison between participants. The scoring method of the Temporal Judgment was altered with the use of percentiles[1]. Participants were awarded two points if their answer fell within the 33rd and 67th percentile range. One point was awarded when an answer fell within the 16th and 33rd or the 67th and 84th percentile range. Zero points were awarded when an answer's percentile was lower than the 16th percentile or higher than the 84th percentile (Leen, 2019). Consequently, the maximum number of points that could be awarded increased from four to eight points.

Furthermore, several statistical analyses were run. Independent samples *t* tests with bootstrapping were used to examine the possible influence of ADHD on performance scores of the BADS, in order to decide whether to include or exclude participants with ADHD from further data analyses. Multiple regression analyses were used to examine possible influences of age, education, and gender on performance scores of the BADS. The new and former norm data were compared with the use of one sample *t* tests. Possible time differences between the age groups were explored with the use of multiple one-way ANOVAs.

Results

Participants

A total of 122 participants met the inclusion criteria and were included in this study. See table 1 below for descriptive statistics and demographical data of the sample. Before data collection, the age ranges were roughly estimated to ensure that the prospective age groups would be similar in size and gender distribution. Ultimately, the sample was divided into four age groups, based on a roughly equal number of participants per age group: Young Adults (ages 18-32), Middle-aged Adults (ages 33-51), Older Adults (ages 52-69), and Elderly (ages 70-91). With the use of boxplots, data outliers for the raw subtest scores were examined for the total sample and for the individual age groups. All outliers were carefully evaluated and possible causes

[1] Data collection for this research was done in collaboration with another student, M. L. Leen, who performed an extensive analysis of the Temporal Judgment questions. For more information on the new scoring method and validation of the pre-existing and newly formulated questions, see Leen (2019). The new scoring method is included in the current study as well, in order to provide adequate new Dutch normative tables.

were considered. One participant was a significant outlier on multiple subtests and was diagnosed with ADHD and PDD-NOS, and therefore entirely removed from the data. One participant quit the experiment after Zoo Map condition 1, but the previous subtests were completed without issue. Therefore, only the participant's data from the first Zoo Map was removed due to an observed lack of motivation. Lastly, a single answer on the Temporal Judgment subtest was removed because of presumed misunderstanding of the question by one participant.

Furthermore, independent samples *t* tests with bootstrapping (Mersenne Twister = 123456) were used to examine group differences on the test scores between participants with ADHD (*N* = 7) and without ADHD (*N* = 114). There were no significant differences between the groups, except for the Zoo Map raw score and Zoo Map profile score. Surprisingly, the participants with ADHD obtained better scores (*M* = 15.29, *SD* = 1.89) on the Zoo Map raw score than the non-ADHD participants (*M* = 12.32, *SD* = 3.98), $t(9.8) = -3.66$, $p = .005$, two-tailed, $d = 0.76$. The same difference was found for the Zoo Map profile score, where participants with ADHD also obtained better scores (*M* = 3.71, *SD* = 0.49) than the non-ADHD participants (*M* = 2.82, *SD* = 1.14), $t(10.8) = -4.19$, $p = .002$, two-tailed, $d = -0.80$. Since there were only significant differences found (in an unexpected direction) for the Zoo Map, and to preserve the approximate number of 30 participants per age group, participants with ADHD were not excluded from further analyses.

Table 1
Descriptive Statistics and Demographical Data of the Sample

Group	***N***	**Male**		**Female**		**Age (years)**			**Education** [a]	
		N	(%)	*N*	(%)	*M*	*SD*	Range	Median	IQR
Total sample	121	57	(47.1)	64	(52.9)	50.83	20.10	18-91	6	(5-6)
Young Adults	31	16	(51.6)	15	(48.4)	24.77	4.06	18-32	6	(5-6)
Middle-aged Adults	29	14	(48.3)	15	(51.7)	42.79	5.53	33-51	6	(5-6)
Older Adults	31	13	(41.9)	18	(58.1)	59.03	5.12	52-69	6	(6-7)
Elderly	30	14	(46.7)	16	(53.3)	77.03	5.08	70-91	5	(4-6)

[a] According to Verhage coding of educational levels (1964)

Norm study

First, multiple regression analyses were executed to see whether the raw subtest scores of the total sample were influenced by age, gender, or education level. Upon careful examination, the choice has been made to exclude the NLV from all analyses. Pearson's correlation demonstrates that the IQ estimations from the NLV still appeared to be significantly correlated with age ($r(119) = .55$, $p < .001$), even after the age adjustments included in the NLV scoring method.

Results of the multiple regression analyses show that all subtests were significantly influenced by age, with the exception of the Temporal Judgment. Furthermore, there were significant effects of gender on the Zoo Map and of education level on the Modified Six Elements. Subsequently, the effects of age, gender, or education on the different age groups were investigated using multiple regression. The results show that the significant effects of age on the six subtest scores and education level on the Modified Six Elements did not remain. However, two significant influences were found regarding the Zoo Map scores. Education level significantly influenced the test scores of the Middle-aged Adults (p = 0.031) and gender significantly influenced the test scores of the Older Adults (p = 0.021). See Appendix B, table 2 for an overview of the multiple regression coefficients. Secondly, the new answer ranges and corresponding raw scores of the Temporal Judgment subtest were calculated with the earlier described use of percentiles (Leen, 2019).

Subsequently, means and standard deviations of the new profile scores of each subtest were calculated in line with the original scoring method. Significant differences between means of the new and former norm data were examined using one sample t tests. See table 3 for the means and standard deviations of the new norm data and the former norm data (Krabbendam & Kalff, 1999), and the t-test results. Furthermore, the means and standard deviations of the raw scores were computed in order to create the new Dutch normative tables (table 4). Moreover, z scores were calculated for every age group and subtest using the means and standard deviations reported in table 4, and classification tables were created. See Appendix C, tables 5a to 5f for the z scores and corresponding classifications.

Table 3

One Sample T-Tests Between Means of Old and New Normative Data, For All Age Groups Separately

	Old norms: New norms:		Group 1 (15-31 years) Young Adults (18-31 years)				
(Sub)test	**Old norms**		**New norms**				
	M	*SD*	*M*	*SD*	*df*	*t*	Sig.
Overall profile score	19.2	2.4	20.81	2.09	30	4.48	<.001
Rule Shift Cards	3.5	0.9	3.71	0.46	30	2.53	.017
Action Program	3.8	0.4	3.74	0.77	30	-0.42	.679
Key Search	2.8	1.2	3.71	0.60	30	8.61	<.001
Temporal Judgment	2.6	1.1	2.55	0.89	30	-0.33	.749
Zoo Map	2.8	0.9	3.35	0.99	30	3.14	.004
Modified Six Elements	3.6	0.9	3.81	0.40	30	2.86	.008
	Old norms: New norms:		**Group 2 (32-47 years) Middle-aged Adults (32-51 years)**				
(Sub)test	**Old norms**		**New norms**				
	M	*SD*	*M*	*SD*	*df*	*t*	Sig.
Overall profile score	18.7	2.8	19.97	2.15	28	3.18	.004
Rule Shift Cards	3.5	0.9	3.45	0.63	28	-0.44	.663
Action Program [a]	3.7	0.9	4.00	0.00	-	-	-
Key Search	3.2	1.0	3.17	1.20	28	-0.13	.902
Temporal Judgment	2.6	0.9	2.41	1.02	28	-0.99	.333
Zoo Map	2.1	1.3	3.21	1.15	28	5.20	<.001
Modified Six Elements	3.6	0.8	3.72	0.53	28	1.27	.216
	Old norms: New norms:		**Group 3 (48-63 years) Older Adults (52-69 years)**				
(Sub)test	**Old norms**		**New norms**				
	M	*SD*	*M*	*SD*	*df*	*t*	Sig.
Overall profile score	18.4	2.8	20.26	1.79	30	5.79	<.001
Rule Shift Cards	3.3	1.0	3.58	0.56	30	2.77	.010
Action Program	3.9	0.4	3.97	0.18	30	2.10	.044
Key Search	3.0	1.2	3.23	1.09	30	1.16	.256
Temporal Judgment	3.0	1.0	2.68	0.87	30	-2.06	.048
Zoo Map	1.9	1.2	2.94	0.85	30	6.75	<.001
Modified Six Elements	3.4	1.0	3.87	0.34	30	7.70	<.001
	Old norms: New norms:		**Group 4 (64-84 years) Elderly (70-91 years)**				
(Sub)test	**Old norms**		**New norms**				
	M	*SD*	*M*	*SD*	*df*	*t*	Sig.
Overall profile score	15.0	4.2	16.57	3.00	28	3.17	.004
Rule Shift Cards	2.3	1.4	3.00	0.95	29	4.05	<.001
Action Program	3.3	1.0	3.30	1.06	29	0.00	1.000
Key Search	2.4	1.3	2.53	1.20	29	0.61	.546
Temporal Judgment	2.2	1.0	2.53	0.73	29	2.50	.018
Zoo Map	1.1	1.5	2.00	1.04	28	2.60	.015
Modified Six Elements	3.4	0.8	3.38	0.90	28	-0.12	.903

[a] Analysis was not possible due to a *SD* of zero

Note. Statistics of the old normative data were reported with one decimal place (Krabbendam & Kalff, 1999)

Table 4

Means and Standard Deviations of the New Normative Data (BADS Raw Scores)

(Sub)test	Young Adults (18-31 years)		Middle-aged Adults (32-51 years)		Older Adults (52-69 years)		Elderly (70-91 years)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Rule Shift Cards	0.29	0.46	0.97	1.27	0.58	1.15	1.93	2.18
Action Program	4.74	0.77	5.00	0.00	4.97	0.18	4.30	1.06
Key Search test	14.61	2.12	13.00	3.64	13.29	3.11	11.20	3.41
(New) Temporal Judgment [a]	5.00	1.55	5.41	1.52	5.48	1.63	5.10	1.50
Zoo Map	14.23	3.51	13.90	3.79	12.39	3.13	9.52	3.61
Modified Six Elements	5.77	0.50	5.79	0.41	5.84	0.45	5.17	1.23

[a] New Temporal Judgment scores were calculated in line with the new scoring method, where 0-2 points could be awarded instead of 0-1 points

Time limit analyses

All group differences were examined using one-way ANOVAs. For several variables the assumption of normality was violated, in which case the Brown-Forsythe test was used. When groups differed significantly ($p < .05$), post-hoc analyses were executed using Tukey HSD to see which group means specifically differed from one another. First, group differences on several performance variables (raw subtest scores and overall profile score) were examined to see whether the groups performed differently on the BADS. See table 6 for an overview of the ANOVA analyses. Even though all four age groups were included in the ANOVA and subsequent post-hoc analyses, the differences most meaningful – and expectedly most notable – for the time limits, are between Elderly and Young Adults. Therefore, only *p* values of post-hoc analyses between these two age groups are reported in the table. Results show significant differences between the age groups on the overall profile score and all subtest raw scores, except for the Temporal Judgment raw score and Zoo Map condition 2 raw score. Post-hoc analyses show that Elderly generally performed more poorly than the other age groups. More specifically, Elderly scored significantly worse than all other age groups on the Rule Shift Cards raw score, Zoo Map condition 1 and total raw score, Modified Six Elements raw score and overall profile score. Furthermore, Elderly scored significantly lower than Older Adults ($p = .003$) on the Action Program raw score, and significantly lower than Young Adults and Older Adults ($p = .048$) on the Key Search raw score. There were no significant differences between the Young Adults, Middle-aged Adults and Older Adults.

Secondly, group differences on several time variables were examined. See table 7 for an overview of the ANOVA analyses. Again, all age groups have been included in the ANOVA and post-hoc analyses, but since the most meaningful and notable differences are expected to be seen between Elderly and Young Adults, only *p* values of post-hoc analyses between these two age groups are reported in the table. Also, the BADS originally does not wield a time restriction for the Action Program. Nevertheless, the duration of the Action Program task completion was timed and reported during this study for 95.9% of the participants, to additionally examine whether the age groups differed in duration of task completion. Results show significant differences between the age groups on all time variables, with the exception of the Key Search time and Zoo Map condition 2 planning time. Post-hoc analyses revealed that Elderly generally took the longest to complete the different subtests. More specifically, Elderly took significantly longer than all other age groups to complete the Rule Shift Cards condition 2, Action Program, and Zoo Map condition 1 total time and condition 2 planning time. Furthermore, Elderly took significantly longer than Middle-aged Adults ($p = .012$) to plan a route during the Zoo Map condition 1. There were no significant differences between the Young Adults, Middle-aged Adults and Older Adults.

However, even though Elderly generally take longer to complete the subtests than the other age groups, it does not necessarily indicate that Elderly generally violate the time limits and should therefore need a more lenient time limit. Table 8 provides an overview of the numbers of participants in each age group that violated the time limits of the subtests. When looking at the number of violations each group made on the different subtests, it can be seen that the groups do not differ that much regarding the Key Search and Zoo Map condition 2 planning time. However, the Rule Shift Cards condition 2 and Zoo Map condition 2 total time show notable differences between the age groups. For the Rule Shift Cards, 10.3% of the Elderly violated the time limit, while 3.6% of the Middle-aged Adults and none of the Young Adults and Older Adults violated the time limit. The difference between groups was even larger for the Zoo Map condition 2 total time, with 24.1% of the Elderly violating the time limit, in contrast to 6.9% for the Middle-aged Adults, 3.2% for the Young Adults, and none of the Older Adults. These violation results should be taken into account when deciding whether the time limits should be altered for older participants.

Table 6

Differences Between Age Groups for Performance Variables

Variable	**Young Adults (18-31 years)**		**Middle-aged Adults (32-51 years)**		**Older Adults (52-69 years)**		**Elderly (70-91 years)**		**ANOVA**			**Post-hoc** [a b]
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	*p*	η^2	
Rule Shift Cards raw score	0.29	0.46	0.97	1.27	0.58	1.15	1.93	2.18	7.87	<.001**	.170	<.001
Action Program raw score [c]	4.74	0.77			4.97	0.18	4.30	1.06	6.00	.004	.120	.065
Key Search raw score	14.61	2.12	13.00	3.64	13.29	3.11	11.20	3.41	6.14	.001*	.137	<.001
Temporal Judgment raw score [d]	5.00	1.55	5.41	1.52	5.48	1.36	5.10	1.50	0.76	.518	.019	
Zoo Map												
Condition 1 raw score	6.55	2.78	6.10	3.50	4.48	3.01	2.38	2.73	11.58	<.001**	.230	<.001
Condition 2 raw score	7.68	1.35	7.79	0.68	7.90	0.40	7.24	2.05	1.49	.227	.038	
Total raw score	14.23	3.51	13.90	3.79	12.39	3.13	9.52	3.61	11.00	<.001**	.222	<.001
Modified Six Elements raw score	5.77	0.50	5.79	0.41	5.84	0.45	5.17	1.23	5.47	.002**	.127	.009
Total profile score	20.87	2.08	19.97	2.15	20.26	1.79	16.72	2.93	19.65	<.001**	.340	<.001

[a] Post-hoc analyses with Tukey HSD
[b] Only *p* values of (significant) differences between Young Adults and Elderly were reported
[c] Middle-aged Adults were excluded from Action Program analysis due to a *SD* of zero
[d] Temporal Judgment raw score was computed according to the new scoring method
* Elderly performed significantly worse ($p < .05$) than Young Adults and one other age group (Middle-aged Adults or Older Adults)
** Elderly performed significantly worse ($p < .05$) than all other age groups

Table 7

Differences Between Age Groups for Time Variables

Variables (time in seconds)	**Young Adults (18-31 years)**		**Middle-aged Adults (32-51 years)**		**Older Adults (52-69 years)**		**Elderly (70-91 years)**		**ANOVA**			**Post-hoc** [a b]
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	*P*	η^2	
Rule Shift Cards condition 2 time	28.77	5.41	32.18	10.80	34.84	8.85	46.34	16.50	13.71	<.001**	.268	<.001
Action Program time	57.41	40.15	45.34	14.74	52.43	24.43	129.60	75.72	22.55	<.001**	.372	<.001
Key Search time	44.81	23.96	52.52	31.42	37.29	25.11	57.10	55.64	1.76	.158	.043	
Zoo Map												
Condition 1 planning time	66.90	64.99	46.17	45.06	86.71	103.13	125.38	144.50	3.50	.020	.084	.095
Condition 1 total time	132.77	76.64	128.34	68.79	182.42	123.12	267.83	141.38	10.76	<.001**	.218	<.001
Condition 2 planning time	7.87	11.48	8.45	13.97	8.61	18.80	8.93	12.80	0.03	.994	<.001	
Condition 2 total time	58.77	27.34	61.48	37.46	63.39	28.06	95.34	37.09	8.10	<.001**	.173	<.001

[a] Post-hoc analyses with Tukey HSD
[b] Only *p* values of (significant) differences between Young Adults and Elderly were reported
* Elderly took significantly longer ($p < .05$) than Young Adults and one other age group (Middle-aged Adults or Older Adults)
** Elderly took significantly longer ($p < .05$) than all other age groups

Table 8

Number of Time Violations Per Age Group and Subtest

Variables (time in seconds)	**Time limit**	**Young Adults (18-31 years)**		**Middle-aged Adults (32-51 years)**		**Older Adults (52-69 years)**		**Elderly (70-91 years)**	
		M	Violations *N* (%)	*M*	Violations *N* (%)	*M*	Violations *N* (%)	*M*	Violations *N* (%)
Rule Shift Cards condition 2	> 67 sec.	28.77	0 (0.0%)	32.18	1 (3.6%)	34.84	0 (0.0%)	46.34	3 (10.3%)
Key Search	> 95 sec.	44.81	0 (0.0%)	52.52	2 (6.9%)	37.29	2 (6.5%)	57.10	3 (10.0%)
Zoo Map									
Condition 2 planning time	> 15 sec.	7.87	6 (19.4%)	8.45	5 (17.2%)	8.61	4 (12.9%)	8.93	5 (17.2%)
Condition 2 total time	> 123 sec.	58.77	1 (3.2%)	61.48	2 (6.9%)	63.39	0 (0.0%)	95.34	7 (24.1%)

Discussion

This study aimed to resolve several issues regarding the Behavioral Assessment of the Dysexecutive Syndrome (BADS), by collecting new Dutch normative data among healthy participants and by performing an extensive evaluation of the current time limits of several of the subtests. With regard to the norm study, the data was divided into four age groups and the new normative data was compared with the normative data currently in use (Krabbendam & Kalff, 1999), demonstrating several significant differences in favour of the new normative data. Interestingly, it is not a new finding that test scores improve over a multiple-year time period. James Flynn (1987) reported the universal phenomenon that IQ scores increased from 5 to 25 points in a single generation, also known as "the Flynn effect". This effect is also the reason that the WAIS-IV UK norms are slightly harder than the WAIS-III UK norms (Pearson, n.d.) A similar finding was reported by Baxendale (2010) for memory function: scores on tests involving the learning and recall of visual material improved significantly in every age range over a 22-year time period. Additionally, Dodge, Zhu, Lee, Chang, and Ganguli (2013) found a Flynn effect in the cognitive domains of psychomotor speed, executive functions, and verbal ability. The true cause of the Flynn effect has not been determined yet. Flynn (1999) made the distinction between IQ test scores, psychometric intelligence, and real-world intelligence, where the increase in IQ test scores could actually be an increase in psychometric intelligence, instead of real-world intelligence. Other researchers think that the increased test scores in fact do reflect a true rise in general intelligence, but only because of the increase in educational attainment in modern day society (Baker et al., 2015), even though some researchers report to have found no educational effects (Dodge et al., 2013).

In addition to the new normative data, the scoring method of the BADS has been altered by basing the new norms and classification tables of the subtests on the raw scores of the participants, instead of profile scores. Boxes 3 and 4 below illustrate the use of the new normative data with the earlier described clinical cases from boxes 1 and 2 (see introduction). As you can see, the new scoring method provides a solution for the performance discrepancy seen with the old normative data, where the interpretations of patients' test scores did not always match with the observations from clinical neuropsychologists. The new normative data provides a stricter – and possibly more accurate – classification of the patients' test performances than the old normative data did.

A 64-year-old woman made the maximum amount of nine errors on the Rule Shift Cards, thus achieving a raw score of -9. With the old normative data, she could be classified as an 'average performer'. However, with the new normative data (M = 0.58, SD = 1.15, see table 5), a z score of -7.32 can be calculated, which indicates that this patient deviates highly significantly from healthy 64-year-olds. Additionally, in table 5a in Appendix C can be seen that a raw score of 9 and corresponding z score of -7.32 can be classified as 'Impaired' for her age.

Box 3. New interpretation of the clinical case from Box 1.

A 60-year-old woman performed poorly on the Zoo Map subtest and achieved a total raw score of 6. With the old normative data, this could be classified as an 'average performance'. However, with the new normative data (M = 12.39, SD = 3.13, see table 5), a z score of -2.04 could be calculated, which indicates that she deviates more than two standard deviations from the mean. Additionally, in table 5e in Appendix C can be seen that a raw score of 6 and corresponding z score of -2.04 can be classified as 'Impaired' for her age.

Box 4. New interpretation of the clinical case from Box 2.

With regard to the time limits evaluation, it was hypothesized that older participants would need more time to complete several of the BADS subtests than younger participants. This hypothesis can be partially confirmed. It was demonstrated that the Elderly generally needed more time than the three younger age groups to complete several (but not all) of the subtests where a time limit was applied. While the other age groups did not differ significantly from one another, the Elderly took significantly longer to complete the second condition of the Rule Shift Cards, the Action Program, and the total second condition of the Zoo Map. However, no significant differences between age groups were found for the Key Search time and for the planning time of the second Zoo Map condition. The finding that only the oldest group differs from the other three groups while the latter do not differ from one another on tasks of cognitive performance, is in line with findings from other studies (Burda et al., 2014; Burda et al., 2017; Davis, Heun, & Kise, 2014; Plumet et al., 2005). Although these other studies focused on task performance instead of task speed, the age-related decline in cognitive functions can provide an explanation that works for both speed and performance outcomes. Several executive functions are of influence on the speed of performance on the BADS subtests, such as working memory, speed of processing, cognitive flexibility, inhibition, planning, and problem solving – although declines in speed of information processing can in turn account for a proportion of age-related declines in the other cognitive domains (Deary et al., 2009). The reason why only the oldest participants differ from the other age groups, is because the rate at which cognitive functions decline across lifespan, increases from approximately 60 years old. Salthouse (2009, 2010) examined age-related decline of several cognitive functions and found similar trajectories for, among others, reasoning, memory, speed of processing, and for performances on tasks that measure speed of processing, cognitive flexibility, inhibition, working memory, attention, and other executive functions. The trajectories show a steady, yet moderate decline of these cognitive functions across the lifespan, starting from 20 years old, but an increased rate of decline from around 60 years old. Hence, the decline rate from 20 to 60 years old is generally not fast enough to cause significant differences between Young Adults (18-31 years old), Middle-aged Adults (32-51 years old), and Older Adults (52-69 years old), but once the cognitive functions start to decline faster, it can cause significant differences between Elderly (70-91 years old) and the other age groups.

It is not surprising that the groups did not differ in planning time on the second Zoo Map condition, since this condition is also known as the 'low-demand trial' (Chamberlain, 2003; Wilson et al., 1998), meaning that this condition should be rather simple for all ages to complete. Besides, the time limit for planning a route in this condition is 15 seconds, which indicates how little time the participants are expected to be needing. However, it is more unclear why there is a lack of differences between groups for the Key Search time. One would expect younger participants to be significantly faster than older participants, since speed of processing and the ability to think of an effective search strategy decline with age (Charness, 1985; Salthouse, 2009). This decline is also reflected by the significant differences between groups for the raw score of the Key Search. Since this is the first study, to our knowledge, that has focused on the evaluation of the time limits of the BADS, no possible explanations are available in other researches. However, from test observations an explanation could be that it takes the same amount of time to draw out an effective, square-covering search route (e.g. horizontal, vertical, concentric), as it takes to draw out an unstructured, inefficient, square-covering search route. That way, the quality of performance is not necessarily reflected by the time needed to complete the task, hence the lack of significant differences between time for the different age groups.

In addition to performing more slowly, it was also examined whether the Elderly performed more poorly than younger participants. In line with findings from earlier research (Burda et al., 2017; Davis et al., 2014; Krabbendam & Kalff, 1999), no significant differences were found between the Young, Middle-Aged, and Older Adults, while the Elderly generally obtained significantly lower raw scores than the three younger age groups on most of the subtests and the overall profile score of the BADS. However, the raw scores of the Temporal Judgment subtest and second Zoo Map condition showed no significant differences between the age groups. Again, it is no surprise that all participants performed similarly on the second Zoo Map condition, considering the low cognitive demand and the rather straightforward instructions of the task. The finding that age does not appear to affect the estimation ability, which is measured by the Temporal Judgment subtest, is also not an uncommon finding. Several studies demonstrate a low or even a lack of correlation between age and performance scores on the Cognitive Estimation Test, another neuropsychological measure of estimation abilities (Gansler, Varvaris, Swenson, & Schretlen, 2014; Wagner, MacPherson, Parente, & Trentini, 2010). This could also explain why the scores of the Temporal Judgment subtest were not influenced by age to begin with, as shown in Appendix B, table 2.

Seeing as the Elderly perform more slowly (and more poorly) on several of the BADS subtests, the question arises whether there should be separate time limits for different age groups. One might expect a decision to make the time limits for the Elderly more lenient in general, but upon closer examination it appears that the age groups performed differently for each subtest and so each subtest should be carefully evaluated. With regard to the second condition of the Rule Shift Cards, the Key Search, and the planning time of the second Zoo Map condition, the Elderly did not violate the time limit that much more often than the other age groups. Additionally, even though the means and standard deviations of the Elderly were higher than those of the Young Adults, they generally remained way below the time limit cut-off score. Thus, there does not seem to be a need to make the time limits of these three time variables more lenient for older participants. When looking at the time limit of the total time to complete the second Zoo Map condition, results show that this time limit was violated much more frequently by the Elderly (24.1%) than the other age groups (0.0% - 6.9%). On one hand this could mean that there might be a need for a more lenient time limit for the Elderly. On the other hand, the importance of changing the time limit depends on the expected change in false positives: e.g. an Elderly who takes significantly longer than others to complete a task, but still does so within the time limit so the slow performance goes unnoticed. Since the majority of the Elderly still managed to complete the task within the time limit and since the current time limit of 123 seconds is within one *SD* of the Elderly's mean, a (slight) adjustment to the time limit will probably not have a large effect on the classification of the Elderly.

Strikingly though, it was observed that the Young Adults consistently completed all the tasks within less than half of the time given for that subtest, while obtaining the highest means for most of the time variables. They even completed the Rule Shift Cards and second Zoo Map condition 1.6 times faster than the Elderly, with little to no violations of the time limits. Thus, perhaps the notion should not be to create more lenient time limits for Elderly, but to make some of the time limits more strict for the Young Adults. This specifically applies to the second condition of the Rule Shift Cards and the total time limit of the second Zoo Map condition. For example, when a 24-year-old participant needs 50 seconds to complete the second Rule Shift

Cards condition, but does so without errors, the participant would still obtain the highest score. However, the data of this study suggests that 24-year-olds generally should be able to complete this task ultimately within 40 seconds (two *SD*s from the Young Adults mean). Thus, a timing of 50 seconds would be significantly deviant and could even be a sign of a suboptimal processing speed after a traumatic brain injury, which currently would go unnoticed by the clinician (a false positive).

With regard to the other time variables, there does not seem to be a need to make the time limits of the Key Search and the second Zoo Map planning time more strict for Young Adults. Even though there were little or no violations of the time limits by Young Adults and their task completions were twice as fast as the time limits, there were no significant differences found between the age groups. Thus, while the Rule Shift Cards and the second Zoo Map condition appear to be in need of stricter time limits for Young Adults, the time limits for the Key Search and planning time of the second Zoo Map condition seem to be appropriate for the current age groups.

The current study has a few strengths that are worth mentioning. First of all, the sample size of this study allowed for the data to be divided into four groups of roughly 30 participants per group, with relatively equal male-to-female ratios throughout the sample. Secondly, the total sample is characterized by a wide age distribution, ranging from 18 to 91 years old. A third strength of this study is the decision to base the new normative data on the raw scores, instead of profile scores, and the immediate illustration of the use of this new method, which demonstrates to be an improvement on the currently used Dutch normative data. Fourth, the decision to time the Action Program subtest and to include this variable in the time-related analyses, strengthens the finding that the older participants complete the tasks more slowly than younger participants.

Nevertheless, this study is not free of limitations. To start with, some may have noticed that with the new normative data, point deductions for exceeding the time limits have not been applied. There are several reasons for this. First of all, because the new normative data was based on raw scores instead of profile scores, subtracting one penalty point from the raw score when exceeding the time limit would not have the same weight as subtracting one point from the profile score. It would be inadequate to copy this method and apply it to the newly collected normative data, because this could lead to overestimation of a participant's true performance. Improvidently increasing the time penalty by subtracting two, three, or even four points is not the solution either, since that might lead to underestimation of a participant's true performance. Besides, the subtests vary in raw score range and difficulty level, so it is likely that different penalty numbers are needed per subtest. Secondly, this study is advocating for the alteration of some of the time limits by being potentially more strict for Young Adults, and therefore not comfortable with applying the existing time limits to the new normative data. Before determining the duration of the time limits and penalty numbers, more extensive research is needed to examine how frequently healthy and clinical participants violate the different time limits, as well as examining how long these different groups take to complete the different subtests.

Another limitation of this study is that the new method of the normative data does not allow for the calculation of an overall performance score, such as an “overall profile score” of the current BADS method. However, it appears that only few clinicians administer all six subtests of the BADS when diagnostically assessing a patient. When putting together a test battery for a diagnostic assessment, it is important for clinicians to keep the energy level of the patient in mind and to try to prevent possible exhaustion of the patient, e.g. by only administering the necessary tests and questionnaires. Therefore, as Chamberlain (2003) also advises, clinicians prefer to select a few subtests of the BADS that are meaningful for their hypothesis, and administer those in addition to other tests of executive functions, such as the Wisconsin Card Sorting Test, the Stroop Test, or the Trail Making Test. Thus, the

A third limitation is the sample size of the current study. Even though the sample of 121 participants and roughly 30 participants per age group is a good start, even more participants are needed to create more representative norm data. For example, all participants in the Middle-aged Adults group obtained the maximum score on the Action Program, which made it impossible to include this group in several analyses due to a variance of zero. Capitani (1997) explains that when a sample in a normative study is divided into groups (based on age, education, or another factor), the subgroups should contain a minimum of 90 participants each. This way, one can be 95% confident that at least 95% of the participants in a subgroup vary between the best and worst scores of the total sample. Thus, since the current sample is divided into four age groups, a minimum sample size of 360 participants is needed.

A fourth limitation is that possible brain injuries or disorders, cognitive impairments, and psychopathologies were ruled out based on simply asking participants whether any of these issues were applicable to them. Especially for older participants, it could be difficult to judge their own cognitive abilities. A more valid way of examining possible cognitive defects could have been the addition of a test of cognitive impairment, such as the Mini Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCa), or the Six Item Cognitive

Impairment Test (6CIT). Administering one of these tests in addition to the BADS would add only a few minutes to the entire session and would be a valuable contribution to the results.

Several suggestions for future research arise from these limitations. First and foremost, a larger sample of healthy participants is needed to resolve the lack of variance in the Action Program performances of Middle-Aged Adults, and to examine the duration of the time limits and penalty sizes for the different age groups. Secondly, data from a large clinical sample should be collected in order to make comparisons with healthy participants, and to determine the definite time limits and the weight of violating these time limits. Thirdly, research should focus on administering the new normative data and compare the new classifications with other tests of executive function, in order to further validate the use of the raw scores instead of the current use of profile scores. Finally, the current data does not indicate whether the Elderly need more time because of an age-related decline in executive functions, or a decline in speed of processing. Future research should include a measure of processing speed in order to determine the true cause of the slower performances by older participants.

In sum, this study provides preliminary new Dutch normative data for the BADS, together with an extensive evaluation of the time limits currently in use. Basing the new normative data on raw scores instead of profile scores appears to be a valuable and promising adjustment, resulting in more reliable classifications of performances. Furthermore, this study provides insight in the speed and task performances across different age groups, and proposes that a stricter time limit should be implemented for younger participants. Although this study takes the first new steps in the right direction, a lot of additional research is needed before the new Dutch normative data can be used in clinical practice.

Acknowledgements

The author would like to thank Dr. Carla Ruis for the excellent supervision, patience, and understanding character during the last two years. Dr. Carla Ruis was always willing to advise, think along, answer questions, and provide feedback. Furthermore, the author would like to thank Maran Leen for the smooth collaboration, all the second opinions, and good communication. It was truly an enjoyable time working together.

References

Baker, D. P., Eslinger, P. J., Benavides, M., Peters, E., Dieckmann, N. F., & Leon, J. (2015). The cognitive impact of the education revolution: A possible cause of the Flynn effect on population IQ. *Intelligence, 49*, 144–158. Doi: 10.1016/j.intell.2015.01.003

Baxendale, S. (2010). The Flynn effect and memory function. *Journal of Clinical and Experimental Neuropsychology, 32*(7), 699–703. Doi:10.1080/13803390903493515

Boelen, D. H., Spikman, J. M., Rietveld, A. C., & Fasotti, L. (2009). Executive dysfunction in chronic brain-injured patients: Assessment in outpatient rehabilitation. *Neuropsychological Rehabilitation*, *19*(5), 625-644. Doi: 10.1080/09602010802613853

Burda, A., Baker, K., Burns, A., Butler, C., Clark, L., Cook, W., … Young, M. (2014). Performance of young and middle-aged adults on cognitive tests. *Journal of Medical Speech-Language Pathology, 21*, 393–404.

Burda, A. N., Andersen, E., Berryman, M., Heun, M., King, C., & Kise, T. (2017). Performance of young, middle-aged, and older adults on tests of executive function. *Canadian Journal of Speech-Language Pathology and Audiology, 41*(3), 277-286.

Burgess, P. W., & Stuss, D. T. (2017). Fifty years of prefrontal cortex research: Impact on assessment. *Journal of the International Neuropsychological Society*, *23*(9-10), 755-767. Doi: 10.1017/S1355617717000704

Capitani, E. (1997). Normative data and neuropsychological assessment. Common problems in clinical practice and research. *Neuropsychological Rehabilitation, 7*(4), 295-309. Doi: 10.1080/713755543

Chamberlain, E. (2003). Behavioural assessment of the dysexecutive syndrome (BADS). *Journal of Occupational Psychology, Employment and Disability*, *5*(2), 33-38.

Charness, N. (1985). Aging and problem solving performance. In N. Charness (Ed.), *Aging and human performance* (pp. 225-259). Chichester, U.K.: John Wiley and Sons.

Crawford, J. R., & Henry, J. D. (2005). Assessment of executive dysfunction. In P. W. Halligan & D. T. Wade (Eds.), *Effectiveness of rehabilitation for cognitive deficits* (pp. 233-245). New York, NY, US: Oxford University Press. Doi: 10.1093/acprof:oso/9780198526544.003.0019

Davis, C., Heun, M., & Kise, T. (2014). *Executive function performance in young, middle-aged, and older adults* (Unpublished research report). University of Northern Iowa: Iowa.

Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., ... & Starr, J. M. (2009). Age-associated cognitive decline. *British Medical Bulletin, 92*(1), 135-152. Doi: 10.1093/bmb/ldp033

Dodge, H. H., Zhu, J., Lee, C.-W., Chang, C.-C. H., & Ganguli, M. (2013). Cohort effects in age-associated cognitive trajectories. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 69*(6), 687–694. Doi:10.1093/gerona/glt181

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*(2), 171-191. Doi: 10.1037/h0090408

Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist, 54*(1), 5–20. Doi: 10.1037/0003-066X.54.1.5

Gansler, D. A., Varvaris, M., Swenson, L., & Schretlen, D. J. (2014). Cognitive estimation and its assessment. *Journal of Clinical and Experimental Neuropsychology, 36*(6), 559-568. Doi: 10.1080/13803395.2014.915933

Ghawami, H., Sadeghi, S., Raghibi, M., & Rahimi-Movaghar, V. (2017). Executive functioning of complicated-mild to moderate traumatic brain injury patients with frontal contusions. *Applied Neuropsychology: Adult, 24*(4), 299-307. Doi: 10.1080/23279095.2016.1157078

Krabbendam, L., & Kalff, A. C. (1999). *Supplement bij de Nederlandse versie van de Behavioural Assessment of the Dysexecutive Syndrome (BADS).* Lisse: Swets & Zeitlinger B.V.

Leen, M. L. (2019). *The development of new temporal judgement questions and gathering new normative data for the Behavioural Assessment of the Executive Syndrome (BADS).* (Unpublished master's thesis). Utrecht University, Utrecht, The Netherlands.

Pearson (n.d.). Wechsler Adult Intelligence Scale - Fourth UK Edition (WAIS-IV UK) - Frequently Asked Questions. Retrieved on June 8th 2020, from https://www.pearsonclinical.co.uk/Psychology/AdultCognitionNeuropsychologyandLanguage/AdultGeneralAbilities/WechslerAdultIntelligenceScale-FourthUKEdition(WAIS-IVUK)/ForThisProduct/FrequentlyAskedQuestions.aspx

Plumet, J., Gil, R., & Gaonac'h, D. (2005). Neuropsychological assessment of executive functions in women: Effects of age and education. *Neuropsychology, 19*(5), 566. Doi: 10.1037/0894-4105.19.5.566

Raz, N., & Rodrigue, K. M. (2006). Differential aging of the brain: Patterns, cognitive correlates and modifiers. *Neuroscience & Biobehavioral Reviews, 30*(6), 730-748. Doi: 10.1016/j.neubiorev.2006.07.001

Salthouse, T. A. (2009). Decomposing age correlations on neuropsychological and cognitive variables. *Journal of the International Neuropsychological Society, 15*(5), 650-661. Doi: 10.1017/S1355617709990385

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society*, *16*(05), 754–760. Doi:10.1017/s1355617710000706

Schmand, B., Lindeboom, J., & Van Harskamp, F. (1992). NLV: Nederlandse leestest voor volwassenen: Handleiding. *CoTAN*, 484-485.

Wagner, G. P., MacPherson, S. E., Parente, M. A. M. P., & Trentini, C. M. (2011). Cognitive estimation abilities in healthy and clinical populations: The use of the Cognitive Estimation Test. *Neurological sciences, 32*(2), 203-210. Doi: 10.1007/s10072-010-0449-4

Waters, A. B., Sawyer, K. S., & Gansler, D. A. (2018). On the impact of interhemispheric white matter: Age, executive functioning, and dedifferentiation in the frontal lobes. *International Journal of Geriatric Psychiatry, 33*(9), 1271-1279. Doi: 10.1002/gps. 4924

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). *Behavioural assessment of the dysexecutive syndrome (BADS)*. London, UK: Thames Valley Test Company.

Appendix A
Description of the BADS Subtests

Rule Shift Cards

During this test participants are presented with a booklet of 21 red and black playing cards. The experimenter turns over the cards one by one and the participants have to answer yes or no according to a specific rule. There are two conditions. During the first condition, the following rule applies: say "yes" to red cards, and say "no" to black cards. During the second condition, participants have to abandon the first rule and adopt a new one: say "yes" when a card has the same colour as the previous card, if not, say "no". This test measures the ability of remembering and adequately switching between the rules, and the ability to remember the colour of the previous card (Wilson et al., 1996).

Action Program

This subtest requires participants to come up with a five-step plan to solve a problem. Participants are presented with a small, transparent container, a tall tube with a cork inside, an L-shaped metal rod, another tube, and a small screw cap. The container is filled with water and closed by a removable lid with a small hole in the centre. The participants are instructed to extract the cork from the tall tube, without moving the tall tube and the container, and without touching the removable lid with their fingers (Wilson et al., 1996).

Key Search

Participants are presented with an A4 size paper on which a square is printed. They are instructed to imagine that the square is a large field where they have lost their keys. With pen they have to show the route they would walk in order to find their keys. This subtest is an analogue of an everyday problem and it measures the ability to come up with an effective search strategy and the ability to evaluate your own performance (Wilson et al., 1996).

Temporal Judgment

This subtests consists of four short questions about events that generally last several seconds to several years. Participants are not expected to know the exact answer to these questions – they are instructed to make a reasonable guess. It is not clearly stated in the manual what this subtest assesses. However, it seems to measure judgment and abstract thinking based on common knowledge (Chamberlain, 2003).

Zoo Map

This test consists of two conditions. During the first condition, participants are instructed to plan a route and visit six specific locations on a zoo map. A number of paths on the map are spotted – these can be walked multiple times. The remaining white paths can only be walked once. This first condition provides little external structure, as opposed to the second condition: participants are presented with the same zoo map, but this time they have to follow a step-by-step instruction to visit the six locations. This test allows for comparison between one's ability to plan in a situation with limited external structure and one's ability to follow a given strategy in a situation with high external structure (Wilson et al., 1996).

Modified Six Elements

This is a test of time-management. Over a ten minute period, participants are instructed to work on three different tasks (storytelling, picture naming, and arithmetic). Each task is divided into two parts: A and B. Participants are instructed to try to work on each individual task in the following ten minutes. However, participants have to follow one rule: they are not allowed to work consecutively on two parts of the same task, such as picture naming A and picture naming B. During this subtest it is not important how well participants perform on the several tasks – what matters is how well participants can plan and organise this task (Wilson et al., 1996).

Appendix B

Multiple Regression Tables

Table 2

Influences of Age, Education Level, and Gender on the Raw Subtest Scores For Each Age Group and Subtest Separately

Rule Shift Card

	Young Adults (18-31 years)			Middle-aged Adults (32-51 years)			Older Adults (52-69 years)			Elderly (70-91 years)			Total group		
	B	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Age	-.02	.02	-.19	.07	.04	.29	.05	.04	.21	.12	.08	.29	.03	.01	.36***
Education	-.12	.19	-.12	.33	.51	.13	-1.18	.59	-.38	-.65	.83	-.15	-.53	.27	-.17
Gender	-.17	.18	-.19	-.06	.50	-.02	.02	.42	.01	-.10	.83	-.02	-.07	.26	-.02

Action Program

	Young Adults (18-31 years)			Middle-aged Adults (32-51 years) [a]			Older Adults (52-69 years)			Elderly (70-91 years)			Total group		
	B	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Age	-.01	.04	-.04	-	-	-	-.01	.01	-.32	-.02	.04	-.11	-.01	.003	-.21*
Education	-.13	.31	-.08	-	-	-	-.004	.09	-.01	.02	.42	.01	.06	.13	.04
Gender	-.38	.29	-.25	-	-	-	-.05	.07	-.15	-.33	.42	-.16	-.18	.13	-.13

Key Search

	Young Adults (18-31 years)			Middle-aged Adults (32-51 years)			Older Adults (52-69 years)			Elderly (70-91 years)			Total group		
	B	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Age	.13	.10	.25	-.07	.13	-.11	-.13	.12	-.22	-.06	.13	-.09	-.07	.01	-.34***
Education	-.12	.86	-.03	1.86	1.48	.25	.87	1.66	.11	-1.64	1.30	-.24	.36	.60	.05
Gender	.54	.81	.13	1.93	1.43	.27	-1.05	1.17	-.17	-1.96	1.29	-.29	-.21	.57	-.03

Temporal Judgment [b]

	Young Adults (18-31 years)			Middle-aged Adults (32-51 years)			Older Adults (52-69 years)			Elderly (70-91 years)			Total group		
	B	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Age	.02	.04	.08	.04	.03	.22	.02	.03	.09	-.03	.028	-.22	.002	.004	.04
Education	-.12	.37	-.06	-.38	.40	-.19	.26	.48	.11	-.09	.29	-.06	-.04	.17	-.02
Gender	.15	.35	.09	.43	.39	.22	.01	.34	.004	-.18	.29	-.12	.10	.16	.06

Zoo Map

	Young Adults (18-31 years)			Middle-aged Adults (32-51 years)			Older Adults (52-69 years)			Elderly (70-91 years)			Total group		
	B	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Age	.08	.17	.10	-.14	.12	-.21	.03	.11	.06	-.23	.13	-.33	-.08	.02	-.42***
Education	.04	1.46	.01	3.26	1.42	.42*	.81	1.54	.10	-.93	1.31	-.13	1.28	.66	.16
Gender	.55	1.38	.08	-.50	1.38	-.07	-2.67	1.09	-.43*	-2.24	1.30	-.31	-1.35	.63	-.17*

Modified Six Elements

	Young Adults (18-31 years)			Middle-aged Adults (32-51 years)			Older Adults (52-69 years)			Elderly (70-91 years)			Total group		
	B	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Age	.03	.02	.25	-.01	.02	-.12	-.01	.02	-.15	-.06	.04	-.25	-.01	.003	-.24**
Education	-.18	.20	-.18	.13	.17	.15	.32	.24	.26	.86	.45	.35	.38	.14	.24**
Gender	-.001	.19	-.001	.05	.17	.06	-.10	.17	-.11	-.09	.44	-.04	-.09	.13	-.06

**p* <.05. ***p* < .01. ****p* < .001.

[a] All participants obtained the maximum score, leaving analysis impossible

[b] Scored according to the original scoring method

Appendix C

Classification Tables of the New Dutch Normative Data

Table 5a

Z scores and Corresponding Classifications of the Rule Shift Cards Subtest (No Time Limits Applied)

Subtest score [a]	Young Adults (18-31 years)	Middle-aged Adults (32-51 years)	Older Adults (52-69 years)	Elderly (70-91 years)	Classification
9+	-18.93	-6.32	-7.32	-3.24	
8	-16.76	-5.54	-6.45	-2.78	
7	-14.59	-4.75	-5.58	-2.33	Impaired
6	-12.41	-3.96	-4.71	-1.87	
5	-10.24	-3.17	-3.84	-1.41	Borderline
4	-8.06	-2.39	-2.97	-0.95	Low Average
3	-5.89	-1.60	-2.10	-0.49	
2	-3.72	-0.81	-1.23	-0.03	
1	-1.54	-0.02	-0.37	0.43	Average
0	0.63	0.76	0.50	0.88	High Average

[a] The subtest score is the number of errors during the second condition

Table 5b

Z scores and Corresponding Classifications of the Action Program Subtest

Subtest score	Young Adults (18-31 years)	Middle-aged Adults (32-51 years) [a]	Older Adults (52-69 years)	Elderly (70-91 years)	Classification
0	-6.16	-	-27.61	-4.06	
1	-4.86	-	-22.06	-3.11	
2	-3.56	-	-16.50	-2.17	Impaired
3	-2.26	-	-10.94	-1.23	Low average
4	-0.96	-	-5.39	-0.28	
5	0.33	-	0.17	0.66	Average

[a] All participants obtained the maximum score, so *z* scores could not be computed due to a *SD* of zero

Table 5c

Z scores and Corresponding Classifications of the Key Search Subtest (No Time Limits Applied)

Subtest score	Young Adults (18-31 years)	Middle-aged Adults (32-51 years)	Older Adults (52-69 years)	Elderly (70-91 years)	Classification
0	-6.89	-3.57	-4.27	-3.28	
1	-6.42	-3.30	-3.95	-2.99	
2	-5.95	-3.02	-3.63	-2.70	
3	-5.48	-2.75	-3.31	-2.40	
4	-5.00	-2.47	-2.99	-2.11	Impaired
5	-4.53	-2.20	-2.67	-1.82	
6	-4.06	-1.92	-2.34	-1.52	Borderline
7	-3.59	-1.65	-2.02	-1.23	
8	-3.12	-1.37	-1.70	-0.94	Low average
9	-2.65	-1.10	-1.38	-0.65	
10	-2.17	-0.82	-1.06	-0.35	
11	-1.70	-0.55	-0.74	-0.06	
12	-1.23	-0.27	-0.41	0.23	
13	-0.76	0.00	-0.09	0.53	Average
14	-0.29	0.27	0.23	0.82	
15	0.18	0.55	0.55	1.11	High Average
16	0.66	0.82	0.87	1.41	High

Table 5d

Z scores and Corresponding Classifications of the Temporal Judgment Subtest, According to the New Scoring Method

Subtest score	Young Adults (18-31 years)	Middle-aged Adults (32-51 years)	Older Adults (52-69 years)	Elderly (70-91 years)	Classification
0	-3.23	-3.56	-3.36	-3.40	
1	-2.58	-2.90	-2.75	-2.73	
2	-1.94	-2.24	-2.13	-2.07	Impaired
3	-1.30	-1.59	-1.52	-1.40	Borderline
4	-0.65	-0.93	-0.90	-0.73	Low average
5	0.00	-0.27	-0.29	-0.07	
6	0.65	0.39	0.32	0.60	Average
7	1.29	1.05	0.93	1.27	High average
8	1.94	1.70	1.54	1.93	High

Table 5e

Z scores and Corresponding Classifications of the Zoo Map Subtest (No Time Limits Applied)

Subtest score	Young Adults (18-31 years)	Middle-aged Adults (32-51 years)	Older Adults (52-69 years)	Elderly (70-91 years)	Classification
0	-4.05	-3.67	-3.96	-2.64	
1	-3.77	-3.40	-3.64	-2.36	
2	-3.48	-3.14	-3.32	-2.08	Impaired
3	-3.20	-2.88	-3.00	-1.81	
4	-2.91	-2.61	-2.68	-1.53	Borderline
5	-2.63	-2.35	-2.36	-1.25	
6	-2.34	-2.08	-2.04	-0.98	
7	-2.06	-1.82	-1.72	-0.70	Low average
8	-1.77	-1.56	-1.40	-0.42	
9	-1.49	-1.29	-1.08	-0.14	
10	-1.21	-1.03	-0.76	0.13	
11	-0.92	-0.77	-0.44	0.41	Average
12	-0.64	-0.50	-0.12	0.69	
13	-0.35	-0.24	0.19	0.96	
14	-0.07	0.03	0.51	1.24	High average
15	0.22	0.29	0.83	1.52	
16	0.50	0.55	1.15	1.80	High

Table 5f

Z scores and Corresponding Classifications of the Modified Six Elements Subtest (No Time Limits Applied)

Subtest score	Young Adults (18-31 years)	Middle-aged Adults (32-51 years)	Older Adults (52-69 years)	Elderly (70-91 years)	Classification
0	-11.54	-14.12	-12.98	-4.20	
1	-9.54	-11.68	-10.76	-3.39	
2	-7.54	-9.24	-8.53	-2.58	Impaired
3	-5.54	-6.80	-6.31	-1.76	Borderline
4	-3.54	-4.37	-4.09	-0.95	Low average
5	-1.54	-1.93	-1.87	-0.14	
6	0.46	0.51	0.36	0.67	Average