Československá psychologie 2023 / ročník LXVII / číslo 1
DOI: 10.51561/cspsych.67.1.30

COMPETING FACTOR MODELS OF COGNITION OF HEALTHY OLDER ADULTS: SUPPORT FOR SUPERAGERS IDENTIFICATION

RADEK HEISSLER[1], HANA GEORGI[2], JIŘÍ KOŽENÝ[1,3], MILOSLAV KOPEČEK[1,3]

[1] National Institute of Mental Health, Klecany
[2] Prague College of Psychosocial Studies, Prague
[3] Department of Psychiatry and Medical Psychology, Third Faculty of Medicine, Charles University, Prague

ABSTRACT

Objectives. Neuropsychological tests employ several cognitive functions to a different extent. Thus, factor structures of various neuropsychological batteries and their analyses show both similarities and discrepancies. The study explores the Czech comprehensive neuropsychological battery for SuperAgers (older people with excellent cognition) from the cross-sectional and longitudinal point of view in respect to its factor structure and its stability over time.

Sample and settings. The study sample consisted of 361 healthy older adults (age 60–94) assessed in years 2012 and 2015 with cognitive tests battery.

Statistical analyses. Data were analyzed with confirmatory factor and invariance analyses over time using multiple competing theory-driven models of cognition based on previous studies consisting of 1–5 factors.

Results. The results show that the best fitting model consists of four factors: verbal memory, attention/working memory, executive functions, and language. The results also suggest that the four factorial structure of cognition in healthy older people was the most stable. This reflects their cognitive functioning and highlights the need to identify the SuperAgers on the basis of performance in multiple cognitive domains. The authors propose that these four domains should be taken into account for identifying SuperAgers and that comparing competing models should be a standard procedure in future studies.

Limitations. The visuospatial or nonverbal memory factors were not represented in our study with relevant tests. Our sample consisted of healthy older adults.

key words:
cognition,
older adults,
superagers,
CFA,
neuropsychological battery

klíčová slova:
kognice,
senioři,
superager,
CFA,
neuropsychologická baterie

1. SUBJECT

In the psychology of aging, cognitive abilities are explored from various perspectives. Known facts of their decline with increasing age or disruption of some of their aspects, findings of disorders such as dementia or Alzheimer's disease are supplemented by more recent trends focusing on the positive side of things related to concepts such as cognitive reserve and maintenance, scaffolding, healthy aging, superaging, etc. (e.g, Nilsson & Lövdén, 2018; Nyberg & Pudas, 2019). SuperAg-

---

Submitted: 15. 3. 2022; R. H., National Institute of Mental Health, Topolová 748, Klecany 250 67, Czech Republic; e-mail: radekheissler@gmail.com
This work was supported by the Czech Science Foundation under grant GA 18-06199S (Cognitive superaging). All the data are available upon reasonable request by contacting the authors, and will be published online after publication.

ers (SA) are usually described as cognitively healthy people 60+ or 80+ years old with memory performance of the 20–30 years younger, and with good non-memory performance (e.g., Červenková et al., 2019; Harrison et al., 2012). In most of the studies, the main criteria of their identification consist of performance in multiple cognitive domains (e.g., memory, executive functions, attention, and language; Dang et al., 2018; Harrison et al., 2012; Saint Martin et al., 2017). However, some studies use only a total score of global cognition screening to identify SA (e.g., Chong et al., 2018; Shi et al., 2016).

1.1. Cognitive Factors and Domains

Generally speaking, to differentiate cognitive domains, various neuropsychological tests try to ideally capture and measure one particular cognitive domain or ability (comprehensive overview and details, e.g., Lezak, 2012; Mitrushina, 2005). This can help to identify the main problems and problematic areas of the person concerned, target treatment or training more effectively, or make a differential diagnosis of specific disorders. However, in this context, it is clear from the current research trends (see below) that the reality of tests is often elsewhere as most are multifactorial and point to the interrelated relationships of cognitive abilities among themselves. Additionally, the scores and subscores of various neuropsychological tests are perceived, used, and interpreted inconsistently by researchers in their studies – for example, tests of verbal functions are also used to measure general cognition, processing speed, etc. (as seen e.g., in comparison of factors and tests in studies Hayden et al., 2011; Noh et al., 2010; Siedlecki et al., 2008). This exacerbates the difficulties in comparing results and deriving consequences from data, and as mentioned by Agelink van Rentergem et al. (2020) there is a lack of consensus on which tests belong to which domain because there are many reasonable ways to assign tests to domains. On the other hand, some studies show that analyses of cognitive domains converge and create specific models, such as the Cattell-Horn-Carroll, or the Strauss, Sherman, and Spreen model (see Agelink van Rentergem et al., 2020). However, this in certain sense points to the interdependence of cognitive domains which may be understood in their complexity and not necessarily independently stand alone.

One way to address the issue of these differences between tests and domains (also over time) is through the use of confirmatory factor analysis (CFA) which allows tests from a neuropsychological battery to be summarized into cognitive domains. It evaluates the construct validity of neuropsychological batteries, how well a hypothesized (usually theory-driven) model fits the observed data, and compares it with specific alternative models (Park et al., 2012; Schretlen et al., 2013). Measures of invariance or invariance analyses are used to estimate the stability of factor structures and the differences in a model, whether relations and measurements between latent variables and their manifest indicators are invariant across occasions (Hayden et al., 2011; Siedlecki et al., 2008).

CFA has been used to test the measurement invariance of neuropsychological batteries in studies to characterize how well these compilations of tests measure cognition in older adults (e.g., Dowling et al., 2010; Hayden et al., 2011; Park et al., 2012; Tuokko et al., 2009). In studies concerning the distribution of cognition, a more detailed examination reveals some inconsistencies and ambiguity regarding the interpretation, categorization, and a number of cognitive domains. For example, studies of healthy older people, people with suspected dementia, mild cognitive impairment, or Alzheimer's disease used various models and combinations of one to seven cognitive factors or domains. Their overview can be seen in Table 1.

Table 1 Overview of the selected factor models of cognition

Nr. of factors	Source	Cognitive domains							
One	Kanne et al., 1998	Global cognition							
Two	Noh et al., 2010	Verbal memory	Attention, language, and executive functions						
Four	Hayden et al., 2011	Memory	Attention	Language	Executive functions				
Five	Siedlecki et al., 2008	Memory	Speed and attention	Language	Reasoning	Visuo-spatial functions			
	Weintraub et al., 2009	Memory	Attention	Language	Executive functions	Processing speed			
	Park et al., 2012	Memory	Attention	Language	Executive functions and processing speed	Visuo-spatial functions			
	Agelink van Rentergem et al., 2020	Long-term memory encoding and retrieval	Processing speed	Word fluency	Working memory	Acquired knowledge or crystallized ability			
Six	Siedlecki et al., 2008	Memory	Attention	Language	Reasoning	Visuo-spatial functions	Speed		
	Loewenstein et al., 2001	Memory	Visuo-spatial functions	Language	Executive functions	Instrumental activities of daily living	Activities of daily living		
Seven	Loewenstein et al., 2001	Memory	Attention	Language	Problem-solving	Functional abilities	Orientation	Praxis	

During the process of categorizing the cognitive domains, various procedures are applied, e.g., gradual merging and reducing the number of domains as shown e.g., in Noh et al. (2010). Also, as seen e.g., in Schretlen et al. (2013) it is possible to analyze multiple cognitive models with a number of domains and combinations of measures (e.g., analyzing one model including a single memory domain, and another model with a memory domain divided into verbal and visual).

As is evident from those and other, even meta-analytic, studies (e.g., Agelink van Rentergem et al., 2020; Aschwanden, et al., 2017; Hoogendam et al., 2014; Murman, 2015; Zaninotto et al., 2018) the most commonly employed cognitive domains are memory, language skills, attention or speed of processing, and executive functions. Sometimes, other more differentiated factors are added, related, e.g., to visual or verbal cognitive functions, reasoning, working memory, or variants and different combinations of abilities (e.g., one factor of attention/processing speed/executive function, cf. An et al., 2018; visual/working memory/executive functions, cf. Dowling et al., 2010). Of the cognitive functions, memory, besides processing speed, is probably the most frequently mentioned and studied in connection with old age, because its disturbance is subjectively perceived the most with increasing age (e.g., Markova et al., 2019; Nyberg et al., 2012).

1.2. Neuropsychological Batteries

In this context, various batteries of neuropsychological tests are created and assembled, for which there is an effort to determine in advance which tests focus on which cognitive domain, to ideally cover their entire planned spectrum. These test batteries can thus reduce inconsistency in administration and improve the comparison of results. Concerning the mentioned problems, inconsistency in studies and during test battery creation can be manifested in at least two levels - firstly in determining the cognitive domains themselves, and secondly, in selecting tests that measure these domains (see also Agelink van Rentergem et al., 2020). Such batteries attempt to measure e.g., memory, attention, executive functions, speech, and visuospatial functions (more detailed information and overview, e.g., Lezak, 2012; Věchetová et al., 2018). One of them is a battery consisting of standardized neuropsychological tests called Uniform Data Set (UDS; Weintraub et al., 2009), focusing on testing the cognition of older people, the second version of which (UDS-v2) also has Czech norms (Nikolai et al., 2018). The original study, using the experts' decision, identifies five domains in the UDS-v2: attention, executive functions, processing speed, episodic memory, and language (Weintraub et al., 2009). A newer study of UDS-v2, using factor analysis, showed four factors: memory, attention, executive functions, and language (Hayden et al., 2011). Unfortunately, due to the nature of the tests used in the UDS-v2, it does not contain a factor that could be described as visuospatial, present in the other test batteries. However, this factor has been included in the newer UDS version 3.0 (Kiselica et al., 2020).

1.3. Current Study

In our previous project, the National Normative Study of Cognitive Determinants of Healthy Ageing (NANOK) ongoing between 2012 and 2015, we were inspired by the UDS-v2 battery, to which were added some frequently used neuropsychological tests or counted alternative scores (see below), that are used, e.g., to identify SuperAgers among healthy older adults.

Given the problems with cognitive domains, their processing, the tests used, and the need to reconcile the theoretical knowledge of cognition and research data itself, our work aims to verify whether our test battery covers usually described cognitive domains in healthy older adults from the Czech Republic by analyzing competing models of cognition. Moreover, which combination of tests and factors represented the cognition of healthy older adults from the Czech Republic in the most optimal and stable way, even after several years.

We hypothesize that identification of the important and stable domains should reflect the cognitive functioning of healthy adults, which can lead, among other things, to the more accurate identification of the SA, as they can be perceived as a special subcategory of healthy older adults. More clarity on the number and types of cognitive domains could facilitate test selection, neuropsychological research, and comparison of studies related also to the SA.

2. METHODS

2.1. Participants

The sample was composed of participants in the project National Normative Study of Cognitive Determinants of Healthy Ageing (NANOK; IGA NT 13145) at the National Institute of Mental Health, a successor of Prague Psychiatric Center in the Czech Republic. The assessment and recruitment (through advertising in the local media, doctor's waiting rooms, post offices, senior institutions, or on the web), took place in 12 regions of the Czech Republic to increase representativeness, with the help of 25 trained psychometrists. Ethical approval for this study was obtained on June 29th, 2011 from the local ethics committee of the Prague Psychiatric Center under reference number 64/11, and written informed consent was obtained from all subjects before the study. This study was not preregistered. Detailed anamnestic inclusion criteria and the complete battery of tests and questionnaires are provided by Štěpánková et al. (2015).

Initially, data from 540 examined subjects aged 60–98 years were included for further analysis, with ex-post exclusion criteria: performance below -2 SD than the group average in two cognitive tests or one cognitive test, and at the same time in the Geriatric Depression Scale (GDS-15; Heissler et al., 2020; Sheikh & Yesavage, 1986) or in the Functional Activities Questionnaire (Bezdíček et al., 2011; Pfeffer et al., 1982). Cognitive tests determining the inclusion criteria in the sample were the Trail Making Test (TMT-B; Bezdíček et al., 2012), the Verbal Fluency Test (measured by a composite score from the semantic fluency: Animals and phonemic fluency tests; Nikolai et al., 2015), the Philadelphia Verbal Learning Test czP(r)VLT-12 (composite score - a sum of experiments 1-5 [learning capacity index] and delayed free recall [retention index; Bezdíček et al., 2014]. In the next stage, those respondents who were examined both in 2012 and 2015 were selected (*n* = 423). Individuals who had undergone only an abbreviated version of the assessment and had, therefore, incomplete data were removed (*n* = 22), also those with a missing score in one of the tests (*n* = 15), and those whose performance in the TMT-A, TMT-B, or the Stroop tests were > 3 SD higher than the rest (for example, due to a visual defect, mental exhaustion causing a practical inability to pass the test; *n* = 25). The total sample thus consisted of 361 mentally healthy older people aged 60–94 years, without serious neurological, oncological, or psychiatric illnesses, 205 of them women (56.8 %), with a mean of age 72.89 ± 8.34, and mean 13.29 ± 3.53 of years of education.

2.2. Measures

All participants completed a battery of the following and briefly described neuropsychological tests in the years 2012 and 2015, most of which are also included in the

Uniform Data Set – version 2 (UDS; Nikolai et al., 2018; Weintraub et al., 2009), that aims to cover and operationalize several cognitive dimensions: memory, attention, executive function, psychomotor speed, language.

Trail Making Test (parts A and B; TMT-A and B; Bezdicek et al., 2012; Schretlen et al., 2013) employs psychomotor speed, speed of processing, and visual attention (especially Part A), and executive functions, especially cognitive flexibility and switching between tasks (especially Part B). The basic scores are the number of errors and time to complete each part from which other scores (difference or ratio) are derived, with lower scores suggesting better functioning. Besides the basic scores, we also measured executive functions, with the difference between the total duration of the tests in parts A and B: TMT: B - A.

Prague Stroop Test (PST; Bezdicek et al., 2015) measures attention, working memory, and executive functions, especially inhibition, where lower scores mean better functionality. The score used in this study was the ratio between the time to complete the first (D: ‚dots') and the third (C: ‚colors') parts of the test (PST C / D), which primarily measures the interference rate as part of executive functions.

Boston Naming Test (BNT, its shorter version BNT-30; Harry & Crowe, 2014; Zemanová et al., 2016) measures verbal functions but is also partly related to visual analysis, memory, semantic, and phonological processing (Pupíková, 2016, p. 76). Scores are the number of correctly named images with or without clues and their sums; the more, the better. The score we used to capture language functions was the sum of correctly spontaneously named images + the number of correct answers after providing a semantic clue (0–30 points).

Philadelphia Verbal Learning Test (PVLT; Bezdíček et al., 2014; Libon et al., 1996) measures verbal memory and learning. In our analysis, the delayed free recall of the first list (PVLT 9) score and the sum of successful immediate recalls in trials 1 to 5 (PVLT 1–5) were used to measure verbal memory, with higher scores suggesting better memory functioning.

Logical Memory (LM, Story; Wechsler, 2011) evaluates the total sum of well-recalled story information in immediate and delayed recall parts. In our analyses, immediate and delayed spontaneous recall scores were used to measure verbal memory and learning (0–25 points in each part, more points mean better memory).

Digit Symbol Substitution Test (DSST; Symbols – coding; Wechsler, 2010) focuses on psychomotor speed, executive functions, attention, visual perception, processing speed, or working memory (Jaeger, 2018). The score used in analyses equaled the total number of correctly assigned numbers to symbols in 120 seconds (0–133 points, the more, the better).

Digit Span (DS; Wechsler, 2010) focuses on attention, working, and short-term memory. The scores used in this study are the total number of correctly repeated number series in both parts (higher scores mean better functioning) – repeating series of numbers forwards, then backwards (0–16 in the first, and 0–14 points in the second part), and the number of digits in the longest correctly recalled series (2–9 in the first, and 2–8 in the second part).

Categorical verbal fluency (VF; animals and vegetables; Moms et al., 1989; Nikolai et al., 2015) focuses on language functions but is also related to semantic memory, psychomotor speed, or executive functions. The task was to list as many words from the category of animals and vegetables as possible in one minute. The sum of non-repeated words per minute was used (a higher number suggests better functioning).

2.3. Data analysis

JASP software version 0.11.1 was used for the analyses. The CFA (maximum likelihood estimation), using the raw scores of the scores (similarly as, e.g., Siedlecki et al., 2008) in the years 2012 and 2015, was performed to test several competing models of cognitive abilities and domains based on theory or previously analyzed data. We‘ve used several formerly tested and suggested models of cognitive domains analyzed on various groups, which were analyzed in our sample by CFA. It has been previously suggested that evaluating multiple competing models and their stability over time, not only one a-priori hypothesized model, might provide stronger support for the findings (Park et al., 2012; Schretlen et al., 2013).

Four levels of factorial invariance with progressively stricter constraints were evaluated: 1) configural invariance: the same pattern of fixed and free factor loadings across time; 2) weak factorial (or metric) invariance: invariant factor loadings across time; 3) strong factorial (or scalar) invariance: invariant factor loadings and intercepts across time; and 4) strict factorial (or residual) invariance: invariant factor loadings, intercepts, and unique factor variances across time (Widaman et al., 2010). This enabled their mutual comparison and allowed a detailed understanding of which combination of tests and factors (in line with known underlying cognitive mechanisms) represented and fit our data the best. Invariance analyses found which factors were the most stable in the three-year comparison of healthy older adults. Theoretical models and the factors and scores used to capture them are found in Table 2. The models listed in Table 2 aimed to capture as best as possible all the cognitive mechanisms known from previous research or neuroscientific paradigm (e.g., language, attention, executive functions, memory, and their combination), and try to reflect the lack of consensus.

The first model with five factors (Attention, Processing speed, Executive function, Language, Verbal memory) is based on the cognitive domains and tests selected in the original UDS (Weintraub et al., 2009), our battery differs by adding the Prague Stroop Test interference score to measure executive functions, and PVLT scores instead of the Rey Auditory Verbal Learning Test (RAVLT) to measure verbal memory. This model has two variants, A and B. Model A uses only the basic TMT-B score to measure executive functions (in addition to the PST score); model B uses TMT-B minus A score. Moreover, we have analyzed and replicated the recently studied and recommended Cattell-Horn-Carroll (CHC) model comprising of five factors – Acquired knowledge/Crystallized ability, Long-term memory encoding and retrieval, Working memory, Processing speed, and Word fluency (Agelink van Rentergem et al., 2020).

The models with four factors (Attention, Executive function, Verbal memory, Language) are inspired by the study of older people with and without dementia that, using the CFA, identifies cognitive domains of the UDS neuropsychological battery, which shares some tests with our battery (Hayden et al., 2011). Compared to the previous model, tests from the factor Processing speed are integrated into other factors. Scores of other tests were added to the UDS tests, according to the best theoretically relevant category, but this gave rise to several different options for this model, listed below.

The four-factor model A differs by using the basic scores in TMT and, in particular, the split-loading of TMT-A and DSST scores, which, given their theoretical focus and possible interpretation, are present in both Attention / Working Memory and Executive functions. In contrast, Model B leaves them only in the Executive functions as in the UDS study. Model C uses a TMT differential score in Executive functions, and TMT-A remains in the Attention / Working Memory factor. The last four-factor model D differs from the previous by removing the TMT-A score from the Attention / Work-

Table 2 Overview of factors and used scores

Model	Cognitive domains / Factors	Scores
5 factors A (5A)	Attention	DS-F T, DS-F L, DS-B T, DS-B L
	Executive functions	PST, TMT-B
	Processing speed	TMT-A, DSST
	Verbal memory	PVLT-IM, PVLT-DEL, LM-IM, LM-DEL
	Language	VF-ANI, VF-VEG, BNT-30
5 factors B (5B)	Attention	DS-F T, DS-F L, DS-B T, DS-B L
	Executive functions	PST, TMT
	Processing speed	TMT-A, DSST
	Verbal memory	PVLT-IM, PVLT-DEL, LM-IM, LM-DEL
	Language	VF-ANI, VF-VEG, BNT-30
5 factors CHC	Acquired knowledge	BNT-30, LM-IM, LM-DEL
	Long-term memory	LM-IM, LM-DEL, PVLT-IM, PVLT-DEL
	Working memory	DS-F T, DS-B T, TMT-B
	Processing speed	TMT-A, TMT-B, DSST
	Word fluency	VF-ANI, VF-VEG
4 factors A (4A)	Attention / Working memory	DS-F T, DS-F L, DS-B T, DS-B L, TMT-A, DSST
	Executive functions	PST, TMT-A, TMT-B, DSST
	Verbal memory	PVLT-IM, PVLT-DEL, LM-IM, LM-DEL
	Language	VF-ANI, VF-VEG, BNT-30
4 factors B (4B)	Attention / Working memory	DS-F T, DS-F L, DS-B T, DS-B L
	Executive functions	PST, TMT-A, TMT-B, DSST
	Verbal memory	PVLT-DEL, PVLT-IM, LM-IM, LM-DEL
	Language	VF-ANI, VF-VEG, BNT-30
4 factors C (4C)	Attention / Working memory	DS-F T, DS-F L, DS-B T, DS-B L, TMT-A
	Executive functions	PST, TMT, DSST
	Verbal memory	PVLT-DEL, PVLT-IM, LM-IM, LM-DEL
	Language	VF-ANI, VF-VEG, BNT-30
4 factors D (4D)	Attention / Working memory	DS-F T, DS-F L, DS-B T, DS-B L
	Executive functions	PST, TMT, DSST
	Verbal memory	PVLT-DEL, PVLT-IM, LM-IM, LM-DEL
	Language	VF-ANI, VF-VEG, BNT-30
3 factors A (3A)	Psychomotor speed / Executive functions	PST, TMT-A, TMT-B, DSST, DS-F T, DS-F L, DS-B T, DS-B L
	Verbal memory	PVLT-DEL, PVLT-IM, LM-IM, LM-DEL
	Language	VF-ANI, VF-VEG, BNT-30
3 factors B (3B)	Psychomotor speed / Executive functions	PST, TMT, DSST, DS-F T, DS-F L, DS-B T, DS-B L
	Verbal memory	PVLT-DEL, PVLT-IM, LM-IM, LM-DEL
	Language	VF-ANI, VF-VEG, BNT-30
2 factors A (2A)	Executive functions / Processing speed	PST, TMT-A, TMT-B, DSST, VF-ANI, VF-VEG, BNT-30, DS-F T, DS-F L, DS-B T, DS-B L
	Verbal memory	PVLT-DEL, PVLT-IM, LM-IM, LM-DEL
2 factors B (2B)	Executive functions / Processing speed	PST, TMT, DSST, VF-ANI, VF-VEG, BNT-30, DS-F T, DS-F L, DS-B T, DS-B L
	Verbal memory	PVLT-DEL, PVLT-IM, LM-IM, LM-DEL
1 factor A (1A)	General cognition	PST, TMT-A, TMT-B, DSST, DS-F T, DS-F L, DS-B T, DS-B L, PVLT-DEL, PVLT-IM, LM-IM, LM-DEL, VF-ANI, VF-VEG, BNT-30
1 factor B (1B)	General cognition	PST, TMT, DSST, DS-F T, DS-F L, DS-B T, DS-B L, PVLT-DEL, PVLT-IM, LM-IM, LM-DEL, VF-ANI, VF-VEG, BNT-30

Notes. CHC = Cattell-Horn-Carroll model; DS-F T = Digit Span - Forward (Total trials), DS-F L = Digit Span – Forward (longest sequence); DS-B T = Digit Span - Backward (Total trials); DS-B L = Digit Span - Backward (longest sequence); PST = Prague Stroop Test (seconds in part Colors / Dots); TMT-A = Trail Making Test-Part A (seconds); TMT-B = Trail Making Test-Part B (seconds); TMT = Trail Making Test (seconds in part B – in part A); DSST = Digit Symbol Substitution Test (total items in 120 s); LM-IM = Logical Memory-Immediate recall (total units); LM-DEL = Logical Memory Delayed recall (total units); PVLT-IM = Philadelphia Verbal Learning Test-Immediate recall (sum of trials 1–5); PVLT-DEL = Philadelphia Verbal Learning Test-Delayed recall (total score in trial 9); VF-ANI = Verbal Fluency-Animals (total in 60s); VF-VEG = Verbal Fluency-Vegetables (total in 60s); BNT-30 = Boston Naming Test-30 (spontaneous recall + semantic clue) ing memory factor for its theoretical redundancy, as it was already present in the TMT differential score.

The three-factor model solves the fragmentation of tests and distinguishes itself by combining the Attention / Working memory domain and the Executive functions into one domain called Psychomotor speed / Executive functions. The three-factor models A and B differ only in which of the TMT scores they use - whether the basic scores or differential scores.

The two-factor model combines tests focused on attention, language, and executive functions (collectively referred to as Processing speed; Noh et al., 2010), and appoints the verbal memory tests to the second factor. As above, versions A and B differ only in the TMT scores used.

The last one-factor model was focused on general cognition and has all tests loaded on one factor. It has been mentioned as a more probable and more representative model of cognition of healthy older people who do not have significantly heterogeneous performance compared to individuals with some neurodegenerative disorder (Hayden et al., 2011). Again, versions A and B differ in the TMT scores.

Within the CFA, due to the interconnection of some test scores (differently computed scores of the same test used in the models), we allowed residual covariances in order to improve the fit of the models (similarly as in Hayen et al., 2011; Noh et al., 2010; in particular Digit Span Forward and Backward - total trials and the longest sequence; and Logical memory - immediate and delayed recall scores. If the model used them, also between TMT-A and TMT-B scores. In single-factor models, also between PVLT scores).

The criterion for assessing the suitability and fit of individual models, whether they also correspond to the analyzed data from different perspectives, was the standard combination of indices (so-called fit indices). A good model should meet, e.g., the following criteria: after calculating the value of χ^2 test and degrees of freedom df, is the ratio of χ^2 / df < 3 (but the size of the sample strongly influences this criterion, Matsunaga, 2010); the value of the index Root mean square error of approximation (RMSEA) < .06 or .08; Standardized Root Mean Square Residual (SRMR) < .1; and incremental Comparative Fit Index (CFI) > .9 or better .95 (more detailed information and description of fit indices and various thresholds in Hu & Bentler, 1999; Matsunaga, 2010; Schreiber et al., 2006).

After assessing the fit of the models, invariance analyses were then carried out on the best-fitting models to determine whether the factor structures are the same or different in various aspects between groups (in this case, between participants' performances in 2012 and 2015, similar to Roesch et al., 2013) - to derive the stability of the factor structure. By using the measurement invariations, we would be able to determine which combination of tests and factors represented the cognition of healthy older adults from the Czech Republic in the most optimal and stable way, even after several years.

For this purpose, the configural, metric, scalar, and strict invariances are used, which focus on the equivalence of factors, their organization and structure, loadings, intercepts, residuals, and unique variations, and add various constraints (more detailed description, e.g., Putnick & Bornstein, 2016).

In the subsequent comparison of the significance of changes in invariances are reused aforementioned fit indices and their changes between models. For metric invariance, significant changes for samples with N > 300 are: ΔCFI by .1, ΔRMSEA by .015, and ΔSRMR by .03. For scalar and strict invariance, ΔCFI by .1, ΔRMSEA by .015, and ΔSRMR by .01 are significant (Chen, 2007).

3. RESULTS

An overview of achieved scores in selected tests in both years is in Table 3. The results of the individual CFA models in the first year are shown in Table 4. It shows that in the year 2012, the best-fitting model is the four-factor model D (4D). Factor covariances of this model in the year 2012 are in Table 5, and its diagram with standardized parameter estimates is in Figure 1.

Subsequently, after applying the required RMSEA and CFI criteria, the invariance analyses were conducted on the best models from each factor group. The criteria were met only by the 5A and 4D models, which were examined by measurement invariance in total to illustrate the differences between them and (in)stability over the years.

Results of measurement invariances and their differences are shown in Table 6. Applying the criteria for changes implies that the 4D model fits the best over three years compared to the 5A, and its factor structure is the most stable. The second model partially met the requirements of invariance, but in total, hasn't had as good results as the 4D. The correlations of all scores in the year 2012 are in Table 7.

Table 3 Scores in the years 2012 and 2015

	Year 2012		Year 2015	
	Mean (SD)	Range	Mean (SD)	Range
DS-F T	8.84 (2.17)	4–16	8.6 (2.18)	4–16
DS-F L	5.86 (1.27)	3–9	5.71 (1.22)	3–9
DS-B T	5.8 (2.09)	2–13	5.6 (2.09)	1–13
DS-B L	4.25 (1.29)	1–8	4.2 (1.17)	2–8
PST	2.41 (.74)	1.15–6.83	2.24 (.69)	1.08–6.19
TMT-A	50.56 (21.86)	11–136	49.5 (21.2)	19–146
TMT-B	128.06 (61.67)	38–399	118.57 (55.48)	41–321
TMT	77.94 (49.46)	-1–336	69.07 (44.75)	-3–242
DSST	50.39 (13.26)	16–95	49.18 (13.68)	14–93
LM-IM	10.01 (3.72)	1–21	9.78 (3.83)	1–23
LM-DEL	8.85 (3.5)	1–20	8.39 (4)	0–21
PVLT-IM	38.81 (7.45)	15–56	39.57 (8.46)	14–60
PVLT-DEL	8.22 (2.59)	0–12	8.3 (2.96)	0–12
VF-ANI	21.64 (5.87)	4–40	20.88 (6.41)	7–39
VF-VEG	14.01 (3.22)	6–24	13.22 (3.43)	5–24
BNT-30	26.59 (2.96)	12–30	26.57 (3.31)	9–30

Note. Abbreviations are the same as in Table 3.

Table 4 Fit indices of the models in the year 2012

Factor model	χ^2	df	χ^2/df	RMSEA (90% CI)	SRMR	CFI
5A	251.64	76	3.31	.080 (.069–.091)	.071	.944
5B	258.34	77	3.36	.081 (.070–.092)	.073	.940
CHC	213.51	41	5.21	.108 (.094–.122)	.078	.907
4A	260.78	78	3.34	.081 (.070–.091)	.072	.942
4B	261.37	80	3.27	.079 (.069–.090)	.072	.943
4C	317.76	81	3.92	.090 (.008–.101)	.080	.921
4D	211.77	68	3.11	.077 (.065–.088)	.072	.950
3A	333.36	83	4.02	.091 (.081–.102)	.090	.921
3B	281.33	71	3.96	.091 (.080–.102)	.092	.927
2A	348.76	85	4.10	.093 (.083–.103)	.092	.916
2B	290.68	73	3.98	.091 (.080–.102)	.091	.924
1A	333.80	85	3.93	.090 (.080–.100)	.081	.921
1B	348.65	86	4.05	.092 (.082–.102)	.082	.912

Note. CHC = Cattell-Horn-Carroll model; df = degrees of freedom; RMSEA = Root mean square error of approximation; CI = Confidence interval; SRMR = Standardized Root Mean Square Residual; CFI = Comparative Fit Index

Table 5 Factor covariances of the 4D model in the year 2012

					95% Confidence Interval	
			Estimate	p	Lower	Upper
Verbal memory	↔	Executive functions	-.484	<.001	-.598	-.370
Verbal memory	↔	Language	.623	<.001	.517	.728
Verbal memory	↔	Attention/working memory	.342	<.001	.221	.463
Executive functions	↔	Language	-.881	<.001	-.986	-.776
Executive functions	↔	Attention/working memory	-.509	<.001	-.639	-.379
Language	↔	Attention/working memory	.524	<.001	.392	.657

4. DISCUSSION

This study aimed to analyze several competing factorial models of human cognition based on previous studies and literature via confirmatory factor analysis and measurement invariance, to find and confirm the best and the most stable factor model of our neuropsychological battery for healthy older adults, and apply these results on the identification of SuperAgers. Our battery for cognitive assessment consists of several well-established tests; most of them are included in the UDS neuropsychological battery that is used to measure cognition and its deficits in older adults (Weintraub et al., 2009).

The competing models differed by using various combinations of tests and selected cognitive domains in conjunction with current theories and research. This approach provides better insight than only analyzing the data in advance with exploratory factor analysis or having only one a-priori hypothesized model. Studies usually define beforehand which tests measure specific cognitive domains and don't analyze other

Table 6 Fit indices and invariances of the competing models in the years 2012 and 2015

Factors	Model	χ^2	df	RMSEA (90% CI)	CFI	SRMR	Δ RMSEA	Δ CFI	Δ SRMR
5A	Configural	509.67	152	.081 (.073–.089)	.949	.071	baseline		
	Metric	543.67	167	.079 (.072–.087)	.946	.085	-.002	-.003	+.014
	Scalar	571.73	177	.079 (.071–.086)	.944	.082	0	-.002	-.003
	Strict	682.83	196	.083 (.076–.090)	.931	.084	+.004	-.013	+.002
4D	Configural	409.35	136	.075 (.066–.083)	.957	.074	baseline		
	Metric	440.19	150	.073 (.065–.081)	.955	.086	-.002	-.002	+.012
	Scalar	482.70	160	.075 (.067–.082)	.950	.083	+.002	-.005	-.003
	Strict	584.18	177	.080 (.073–.087)	.936	.086	+.005	-.014	+.003

Notes. df = degrees of freedom; RMSEA = Root mean square error of approximation; CI = Confidence interval; SRMR = Standardized Root Mean Square Residual; CFI = Comparative Fit Index

Table 7 Correlations of tests in year 2012

	LM-IM	LM-DEL	VF-ANI	VF-VEG	TMT-A	TMT-B	DSST	BNT-30	PST	PVLT--IM	PVLT--DEL	DS-F T	DS-B T	DS-F L	DS-B L
LM-IM	—														
LM-DEL	.872***	—													
VF-ANI	.191***	.275***	—												
VF-VEG	.210 ***	.217***	.416***	—											
TMT-A	.031	-.039	-.379***	-.119*	—										
TMT-B	-.111*	-.156**	-.446***	-.238***	.681***	—									
DSST	.284***	.299***	.460***	.322***	-.433***	-.538***	—								
BNT-30	.385 ***	.346***	.347***	.275***	-.098	-.357***	.378***	—							
PST	-.079	-.069	-.085	-.012	.092	.134*	-.114*	-.154**	—						
PVLT-IM	.347***	.360***	.326***	.427***	-.123*	-.194***	.367***	.263***	-.015	—					
PVLT--DEL	.173***	.234***	.335***	.358***	-.193***	-.265***	.341***	.210***	-.054	.738***	—				
DS-F T	.200***	.218***	.256***	.117*	-.239***	-.240***	.259***	.196***	-.071	.155**	.136**	—			
DS-B T	.285***	.284***	.297 ***	.142**	-.195***	-.263***	.275***	.238***	-.126*	.238***	.176***	.515***	—		
DS-F L	.202***	.232***	.295***	.125*	-.235***	-.222***	.302***	.224***	-.086	.202***	.190***	.901***	.526***	—	
DS-B L	.257***	.266***	.307***	.104*	-.184***	-.272***	.301***	.252***	-.134*	.277***	.197***	.433***	.913***	.489***	—

Note. Abbreviations are the same as in Table 3; * p < .05, ** p < .01, *** p < .001

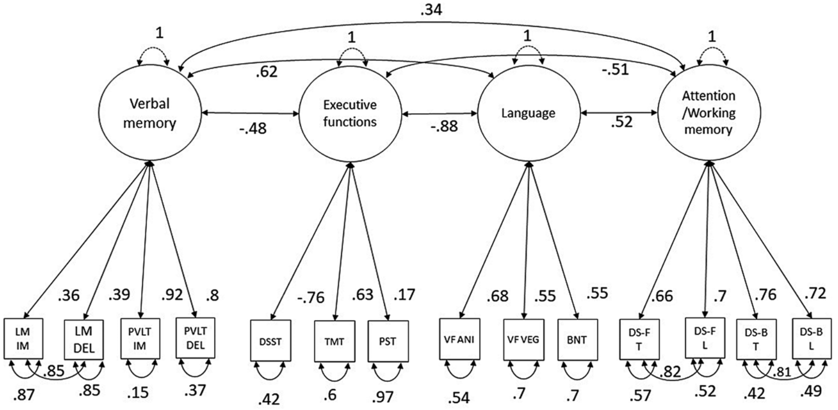

Figure 1 The best-fitting confirmatory factor analysis model 4D with standardized parameter estimates

valid possible combinations, or they skip the process of showing the results of measurement invariance for the different factor solutions and only mention the best fitting model (e.g., Schretlen et al., 2013; Siedlecki et al., 2008). However, some studies have recently used a similar approach as our study (e.g. Agelink van Rentergem et al., 2020). As far as we know, this is the first study that analyzes competing factor models of cognition over time in healthy older adults in the Czech Republic.

Our models had 1 to 5 factors with 2–4 variations based on the tests and scores used that provided different solutions and combinations. After analyzing their fit indices in the year 2012, the models from each factor group that had the best fit requirements were selected for measurement invariance to analyze and compare their stability in the year 2015. However, the differences between the factor solutions were in some cases minimal. We added this (often neglected) step as it illustrates better how (un)stable these factor structures can be over time and how problematic they can be in the long term by failing to interpret the data properly. It can be expected that a model with an initial poor fit would show poor results in invariance analysis, too, but adding these results might provide information and insight into why these models fail to interpret the data and should be abandoned.

We can conclude that the most fitting and plausible model for our neuropsychological battery for older adults and observed data consists of four factors: attention/working memory, executive functions, verbal memory, and language (see Table 2 – model 4D for specific scores used to measure these domains). It is in line with the original factor analysis of the UDS-v2 battery that has also found the four factors model as the most optimal (Hayden et al., 2011). This is not surprising as most of our tests and scores were similar. Our findings are in contrast to models from some previous studies which demonstrated a weaker fit than the model with four factors, despite improving their fit via allowing split-loadings or correlations among various related scores. As shown in Table 4, other factor structures suggested by former studies (e.g., Siedlecki et al., 2008) didn't meet the required fit criteria as well as the four-factor one. It should be noted that the results of those models were not significantly worse and the differences were small, therefore we suggest that they can at least partially represent and interpret our data. The one-factor solution that represented the cognition of healthy older adults according to one study (Hayden et al., 2011) didn't have a better fit than the four-factor one. Similarly, the recently recommended CHC model (Agelink van Rentergem et al., 2020) didn't show a better performance.

The source of these differences can be found in the usage of different tests, scores, and methods (no competing models, only explanatory factor analysis, etc.) in other studies or some specifics of the samples that lead to different factor solutions, as our sample consisted of mentally healthy older people. Notably, some models were based on either experts' decisions or on psychometric properties and statistics which may have led to the disparity. Also, the presence of multiple factors supports the idea of the differentiation of human cognition into various domains even in older healthy adults, suggesting its heterogeneous yet not separate functioning, i.e., the domains being independent to a degree (Harvey, 2019; Pulvermüller et al., 2014). However, this seems to be conflicting with the proposal of Hayden et al. (2011) that the cognition of healthy older adults can be represented by a single factor due to their homogenous performance. We suggest that having multiple corresponding tests and scores to operationalize would lead to satisfactory descriptions, keeping in mind that the cognitive domains and even the tests are interrelated and usually do not capture a precisely single cognitive domain.

This interrelation is also supported by the correlations between our factors and the scores (see Tables 5 and 7), which can all be considered notably large. This is not surprising, as most of the tests are multifactorial and related cognitive domains are considered to be interconnected and influencing each other (Dowling et al., 2010; Siedlecki et al., 2008). However, what needs to be stressed is the fact that merging these factors (a step that might be considered to be logical based on these correlations) hasn't improved the overall fit of the subsequent model, suggesting that differentiating the cognitive domains is appropriate. Moreover, almost all factor-test scores estimates are quite high (see Figure 1), and the tests saturated an appropriate latent domain they were designed to measure based on theoretical background and previous studies. The only lower one is the PST's interference score (.17), which can be caused by its nature (ratio score) and related to its small correlations with other scores (see Table 7). Nonetheless, due to its theoretical relevance and statistical significance, we've included it in the executive functions as the ratio score is preferred in executive functions measurement (Bezdicek et al., 2015).

The most fitting factor structure differs from the original UDS factor analysis by using some added and recalculated scores: the memory domain contains PVLT scores (immediate and delayed recall), the executive functions are measured by the TMT-B minus A score, and the ratio of the Prague Stroop test (colors/dots). It functions as a sort of psychological triangulation because using more tests to measure the same domain should lead to more certain conclusions about its functionality (Jak et al., 2009). By adding these tests and scores into our battery, we were able to create and analyze more (yet still theoretically valid) combinations of tests and factors.

Quite surprising was the fact that in our analyses models using TMT subscore TMT-B minus A showed better results than those with raw TMT-B in several models, and that the best-fitting model doesn't include a TMT-A score at all. A possible interpretation might be that the difference score evaluates executive functions more accurately than the raw TMT-B score by adjusting the influence of visual search and motor speed, measuring cognitive flexibility relatively independently of manual dexterity (Corrigan & Hinkeldey, 1987; Vazzana et al., 2010). The absence of a TMT-A score might be influenced by the fact that related cognitive domains have been saturated enough by other tests (e.g., DSST, Digit Span), and the score itself had no added value in that model even though it is still needed to calculate the TMT difference score. As shown in this study, the selection of appropriate tests and scores to capture and quantify specific domains can be tricky and lead to a bit different results, and authors should choose and think thoroughly and carefully. However, our results showed quite similar, generally good results for all factor structures and models using various scores. Therefore, we suggest that selecting different scores will probably not cause a serious problem in clinical practice as long as the basic cognitive domains are covered with valid methods such as those used in this study.

Our results are related to the SuperAgers (SA) – the cognitively healthy older people (usually aged 80 years and over) with memory performance of the 20–30 years younger. Most of the studies select them using the criteria of performance in multiple domains (scores in PVLT delayed recall, TMT-B, BNT-30, and Animal fluency, or related tests that focus on memory/executive functions/attention/language; Dang et al., 2018; Harrison et al., 2012; Saint Martin et al., 2017). However, some studies use only a score of global cognition screening to select SA (Chong et al., 2018; Shi et al., 2016). Because our sample consists of healthy older adults (and due to their definition, SA can be perceived as a special subcategory of healthy adults), and the most stable and fitting factor solution consisted of four domains, our findings support the idea that the SA should not be selected solely via their performance, e.g., in a single memory or a global cognition screening test, but their performance in multiple cognitive domains should be taken into account. As we have used multiple scores for each cognitive domain, it can be hypothesized that other tests and their scores primarily focused on capturing that particular domain might be used to identify SA (maintaining the requirements of performance as good as younger adults), as seen in some recent studies (Dang et al., 2018; Saint Martin et al., 2017).

Researchers usually use various tests and scores that lead to different factor structures of cognition, which should be a matter of discussion in this field of study. This relates to a limitation of our battery that the suggested four-factor structure captures some basic cognitive functions, but other studies with five or more factors also include, e.g., the visuospatial or nonverbal memory factors (Loewenstein et al., 2001; Park et al., 2012) that were not represented in our study with relevant tests. This was caused by the influence of the UDS 2.0 battery on our battery at that time, which unfortunately lacks these tests. This issue has been addressed in the newest UDS version 3.0 which already includes tests of visuospatial functions and nonverbal memory (Weintraub et al., 2018). It can be, unfortunately, only hypothesized that by adding these tests into our battery, we would have also achieved a valid five or more factors structure as other above-mentioned studies.

This also shows that knowledge in this field of study progresses, tests and batteries change over time, and it should be a good practice when trying to capture global cognitive functioning to cover as many cognitive domains as needed in the particular study as also mentioned by Agelink van Rentergem et al. (2020). A solution might be using a large battery of tests that cover specific cognitive domains from different perspectives via more tests, but its clinical use would be drastically reduced as some neuropsychological tests might be quite lengthy and exhaustive for older adults. It is solved by using brief screening tools in clinical settings that provide some basic information about cognition (De Roeck et al., 2019; Ismail et al., 2010). This trade–off between the complexity and usability of tests and neuropsychological batteries might be one of the major causes of discrepancies among studies and results. However, a more limited number of tests might not possess adequate diagnostic sensitivity and specificity in diagnostic classification (Loewenstein et al., 2001).

Further limitations of this study relate to our sample. Although its size as a whole is adequate for analyses, we haven't analyzed some subsamples divided, e.g., by demographics, as the subsequent samples wouldn't be large enough for valid analyses.

Therefore, larger samples are recommended for future studies that may, e.g., analyze differences in cognitive domains among various demographic groups. Our sample also consists solely of relatively healthy older adults, while former factor studies usually compare groups with different cognitive statuses, e.g., healthy, with mild cognitive impairment, and Alzheimer's disease (e.g., Park et al., 2012; Siedlecki et al., 2008). We have analyzed healthy older adults over time (as done, e.g., in Roesch et al., 2013), as our research is focused on the SuperAgers, who are by definition relatively cognitively intact and with excellent cognition, therefore analyzing people with some cognitive impairment wouldn't fit in this study. Nonetheless, a suggestion for further studies might be comparing whether there are some differences in cognitive domains and their structure between SuperAgers and other people. However, due to the prevalence of SuperAgers (cf. Cervenkova et al., 2019), a much larger sample would be needed.

5. CONCLUSION

Our analyses address the extensive range of neuropsychological test batteries on offer today. Defining latent structure helps to minimize redundancy and increase the reliability of the employed measures. The reported four factors of the latent structure measured by our battery seem to remain stable over several years. Future authors should be aware of which tests measure which cognitive domain in their and other studies, and using standardized neuropsychological batteries is recommended. Due to the complexity and interconnection of human cognition and its models, we suggest that our approach of comparing competing models should be a standard procedure in future studies addressing factor structures of neuropsychological batteries and tests. This study provides the framework, psychometric evaluation, and support of the application of our neuropsychological battery to identify the SuperAgers. Our results might also guide future research designs to maximize the accuracy of neuropsychological assessment of healthy older adults, and the identification of the SuperAgers.

BIBLIOGRAPHY

Agelink van Rentergem, J. A., de Vent, N. R., Schmand, B. A., Murre, J. M. J., Staaks, J. P. C., & Huizenga, H. M. (2020). The factor structure of cognitive functioning in cognitively healthy participants: A meta-analysis and meta-analysis of individual participant data. *Neuropsychology Review*, *30*(1), 51-96. https://doi.org/10.1007/s11065-019-09423-6

An, Y., Feng, L., Zhang, X., Wang, Y., Wang, Y., Tao, L., Lu, Y., Qin, Z., & Xiao, R. (2018). Patterns of cognitive function in middle-aged and elderly Chinese adults–Findings from the EMCOA study. *Alzheimer's Research & Therapy*, *10*(1). https://doi.org/10.1186/s13195-018-0421-8

Aschwanden, D., Martin, M., & Allemand, M. (2017). Cognitive abilities and personality traits in old age across four years: More stability than change. *Journal of Research in Personality*, *70*, 202-213. https://doi.org/10.1016/j.jrp.2017.08.002

Bezdíček, O., Lukavský, J., & Preiss, M. (2011). Validizační studie české verze dotazníku FAQ. *Česká a slovenská neurologie a neurochirurgie*, *74*(1), 36-42.

Bezdicek, O., Motak, L., Axelrod, B. N., Preiss, M., Nikolai, T., Vyhnalek, M., Poreh, A., & Ruzicka, E. (2012). Czech version of the Trail Making Test: Normative data and clinical utility. *Archives of Clinical Neuropsychology*, *27*(8), 906-914. https://doi.org/10.1093/arclin/acs084

Bezdicek, O., Libon, D. J., Stepankova, H., Panenkova, E., Lukavsky, J., Garrett, K. D., Lamar, M., Price, C. C., & Kopecek, M. (2014). Development, validity, and normative data study for the 12-word Philadelphia Verbal Learning Test [czP(r)VLT-12] among older and very old Czech adults. *The Clinical Neuropsychologist*, *28*(7), 1162-1181. https://doi.org/10.1080/13854046.2014.952666

Bezdicek, O., Lukavsky, J., Stepankova, H., Nikolai, T., Axelrod, B. N., Michalec, J., Růžička, E., & Kopecek, M. (2015). The Prague Stroop Test: Normative standards in older Czech adults and discriminative validity for mild cognitive impairment in Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology*, *37*(8), 794-807. https://doi.org/10.1080/13803395.2015.1057106

Corrigan, J. D., & Hinkeldey, N. S. (1987). Relationships between parts A and B of the Trail Making Test. *Journal of Clinical Psychology*, *43*(4), 402-409. https://doi.org/10.1002/1097-4679(198707)43:4<402::AIDJCLP2270430411>3.0.CO;2-E

Červenková, M., Heissler, R., & Kopeček, M. (2019). Stability of memory SuperAgers over 3 years. *PsyCh Journal*, pchj.313. https://doi.org/10.1002/pchj.313

Dang, C., Harrington, K. D., Lim, Y. Y., Ames, D., Hassenstab, J., Laws, S. M., Yassi, N., Hickey, M., Rainey-Smith, S. R., Robertson, J., Rowe, C. C., Sohrabi, H. R., Salvado, O., Weinborn, M., Villemagne, V. L., Masters, C. L., Maruff, P., & AIBL Research Group (2019). Superior memory reduces 8-year risk of mild cognitive impairment and dementia but not amyloid β-associated cognitive decline in older adults. *Archives of Clinical Neuropsychology*, *34*(5), 585-598. https://doi.org/10.1093/arclin/acy078

De Roeck, E. E., De Deyn, P. P., Dierckx, E., & Engelborghs, S. (2019). Brief cognitive screening instruments for early detection of Alzheimer's disease: a systematic review. *Alzheimer's Research & Therapy*, *11*(1). https://doi.org/10.1186/s13195-019-0474-3

Dowling, N. M., Hermann, B., La Rue, A., & Sager, M. A. (2010). Latent structure and factorial invariance of a neuropsychological test battery for the study of preclinical Alzheimer's disease. *Neuropsychology*, *24*(6), 742-756. https://doi.org/10.1037/a0020176

Harrison, T. M., Weintraub, S., Mesulam, M.-M., & Rogalski, E. (2012). Superior memory and higher cortical volumes in unusually successful cognitive aging. *Journal of the International Neuropsychological Society*, *18*(06), 1081-1085. https://doi.org/10.1017/S1355617712000847

Harry, A., & Crowe, S. F. (2014). Is the Boston Naming Test still fit for purpose? *The Clinical Neuropsychologist*, *28*(3), 486-504. https://doi.org/10.1080/13854046.2014.892155

Harvey, P. D. (2019). Domains of cognition and their assessment. *Dialogues in Clinical Neuroscience, 21*(3), 227-237. https://doi.org/10.31887/DCNS.2019.21.3/pharvey

Hayden, K. M., Jones, R. N., Zimmer, C., Plassman, B. L., Browndyke, J. N., Pieper, C., Warren, L. H., & Welsh-Bohmer, K. A. (2011). Factor structure of the National Alzheimer's Coordinating Centers uniform dataset neuropsychological battery: an evaluation of invariance between and within groups over time. *Alzheimer Disease & Associated Disorders*, *25*(2), 128-137. https://doi.org/10.1097/WAD.0b013e3181ffa76d

Heissler, R., Červenková, M., Kopeček, M., Georgi, H. (2020). Geriatrická škála deprese (GDS-15): česká normativní studie [Geriatric Depression Scale (GDS-15): Czech normative study]. *Československá psychologie*, *64*(1), 49-65.

Hoogendam, Y. Y., Hofman, A., van der Geest, J. N., van der Lugt, A., & Ikram, M. A. (2014). Patterns of cognitive function in aging: the Rotterdam study. *European Journal of Epidemiology*, *29*(2), 133-140. https://doi.org/10.1007/s10654-014-9885-4

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Chong, M. S., Tan, C. N., Yew, S., Lim, J. P., Lim, W. S., & Lieu, P. K. (2018). Successful ageing in nonagenarians: Bio-psychosocial factors determining successful ageing in long-lived older adults. *Journal of the American Medical Directors Association*, *19*(1), 90-92. https://doi.org/10.1016/j.jamda.2017.10.015

Ismail, Z., Rajji, T. K., & Shulman, K. I. (2010). Brief cognitive screening instruments: an update. *International Journal of Geriatric Psychiatry*, *25*(2), 111-120. https://doi.org/10.1002/gps.2306

Jaeger, J. (2018). Digit Symbol Substitution Test: The case for sensitivity over specificity in neuropsychological testing. *Journal of Clinical Psychopharmacology*, *38*(5), 513-519. https://doi.org/10.1097/JCP.0000000000000941

Jak, A. J., Bondi, M. W., Delano-Wood, L., Wierenga, C., Corey-Bloom, J., Salmon, D. P., & Delis, D. C. (2009). Quantification of five neuropsychological approaches to defining mild cognitive impairment. *The American Journal of Geriatric Psychiatry*, *17*(5), 368-375. https://doi.org/10.1097/JGP.0b013e31819431d5

Kiselica, A. M., Webber, T. A., & Benge, J. F. (2020). The Uniform Dataset 3.0 neuropsychological battery: Factor structure, invariance testing, and demographically adjusted factor score calculation. *Journal of the International Neuropsychological Society*, *26*(6), 576-586. https://doi.org/10.1017/S135561772000003X

Lezak, M. D. (Ed.) (2012). *Neuropsychological assessment* (5th ed). Oxford University Press.

Libon, D. J., Mattson, R. E., Glosser, G., Kaplan, E., Malamut, B. L., Sands, L. P., Swenson, R., & Cloud, B. S. (1996). A nine–word dementia version of the California Verbal Learning Test. *The Clinical Neuropsychologist*, *10*(3), 237-244. https://doi.org/10.1080/13854049608406686

Loewenstein, D. A., Ownby, R. M. D., Schram, L., Acevedo, A., Rubert, M., & Argelles, T. (2001). An evaluation of the NINCDS-ADRDA neuropsychological criteria for the assessment of Alzheimer's disease: A confirmatory factor analysis of single versus multi-factor models. *Journal of Clinical and Experimental Neuropsychology*, *23*(3), 274-284. https://doi.org/10.1076/jcen.23.3.274.1191

Markova, H., Nikolai, T., Mazancova, A. F., Cechova, K., Sheardova, K., Georgi, H., Kopecek, M., Laczo, J., Hort, J., & Vyhnalek, M. (2019). Differences in subjective cognitive complaints between non-demented older adults from a memory clinic and the community. *Journal of Alzheimer's Disease*, *70*(1), 61-73. https://doi.org/10.3233/JAD-180630

Matsunaga, M. (2010). How to factor-analyze your data right: do's, don'ts, and how-to's. *International Journal of Psychological Research*, *3*(1), 97-110.

Mitrushina, M. N. (Ed.) (2005). *Handbook of normative data for neuropsychological assessment* (2nd ed). Oxford University Press.

Moms, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., Mellits, E. D., & Clark, C. (1989). The consortium to establish a registry for Alzheimer's disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, *39*(9), 1159-1159. https://doi.org/10.1212/WNL.39.9.1159

Murman, D. (2015). The impact of age on cognition. *Seminars in Hearing*, *36*(03), 111-121. https://doi.org/10.1055/s-0035-1555115

Nikolai, T., Stepankova, H., Kopecek, M., Sulc, Z., Vyhnalek, M., & Bezdicek, O. (2018). The Uniform Data Set, Czech version: Normative data in older adults from an international perspective. *Journal of Alzheimer's Disease*, *61*(3), 1233-1240. https://doi.org/10.3233/JAD-170595

Nikolai, T., Štěpánková, H., Michalec, J., Bezdíček, O., Horáková, K., Marková, H., Růžička, E., & Kopeček, M. (2015). Tests of verbal fluency, Czech normative study in older patients. *Česká a slovenská neurologie a neurochirurgie*, *78/111*(3), 292-299. https://doi.org/10.14735/amcsnn2015292

Nilsson, J., & Lövdén, M. (2018). Naming is not explaining: future directions for the "cognitive reserve" and "brain maintenance" theories. *Alzheimer's Research & Therapy*, *10*(1). https://doi.org/10.1186/s13195-018-0365-z

Noh, J., Kim, J.-H., Hong, K. S., Kim, N., Nam, H. J., Lee, D., & Yoon, S. C. (2010). Factor structure of the neurocognitive tests: An application of the confirmative factor analysis in stabilized schizophrenia patients. *Journal of Korean Medical Science*, *25*(2), 276. https://doi.org/10.3346/jkms.2010.25.2.276

Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., & Bäckman, L. (2012). Memory aging and brain maintenance. *Trends in Cognitive Sciences*, *16*(5), 292-305. https://doi.org/10.1016/j.tics.2012.04.005

Nyberg, L., & Pudas, S. (2019). Successful memory aging. *Annual Review of Psychology*, *70*(1), 219-243. https://doi.org/10.1146/annurev-psych-010418-103052

Park, L. Q., Gross, A. L., McLaren, D. G., Pa, J., Johnson, J. K., Mitchell, M., Manly, J. J., & Alzheimer's Disease Neuroimaging Initiative. (2012). Confirmatory factor analysis of the ADNI neuropsychological battery. *Brain Imaging and Behavior*, *6*(4), 528-539. https://doi.org/10.1007/s11682-012-9190-3

Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H., Chance, J. M., & Filos, S. (1982). Measurement of functional activities in older adults in the community. *Journal of Gerontology*, *37*(3), 323-329. https://doi.org/10.1093/geronj/37.3.323

Pulvermüller, F., Garagnani, M., & Wennekers, T. (2014). Thinking in circuits: Toward neurobiological explanation in cognitive neuroscience. *Biological Cybernetics, 108*(5), 573-593.

Pupíková, M. (2016). *Dílčí validizační studie zkrácené verze Bostonského testu pojmenování* (Master's thesis). Masaryk University, Faculty of Arts. https://is.muni.cz/th/xy6hk/

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71-90. https://doi.org/10.1016/j.dr.2016.06.004

Roesch, S. C., Norman, G. J., Merz, E. L., Sallis, J. F., & Patrick, K. (2013). Longitudinal measurement invariance of psychosocial measures in physical activity research: an application to adolescent data: Longitudinal measurement invariance. *Journal of Applied Social Psychology*, *43*(4), 721-729. https://doi.org/10.1111/jasp.12000

Saint Martin, M., Sforza, E., Barthélémy, J. C., Roche, F., Lefèvre, P., Liénard, G., & Thomas-Anterion, C. (2017). Long-lasting active lifestyle and successful cognitive aging in a healthy elderly population: The PROOF cohort. *Revue Neurologique*, *173*(10), 637-644. https://doi.org/10.1016/j.neurol.2017.05.009

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist*, *5*, 165-173.

Shi, W. H., Zhang, H. Y., Zhang, J., Lyu, Y. B., Brasher, M. S., Yin, Z. X., Luo, J. S., Hu, D. S., Fen, L., & Shi, X. M. (2016). The status and associated factors of successful aging among older adults residing in longevity areas in China. *Biomedical and Environmental Sciences: BES*, *29*(5), 347-355. https://doi.org/10.3967/bes2016.045

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323-338. https://doi.org/10.3200/JOER.99.6.323-338

Schretlen, D. J., Peña, J., Aretouli, E., Orue, I., Cascella, N. G., Pearlson, G. D., & Ojeda, N. (2013). Confirmatory factor analysis reveals a latent cognitive structure common to bipolar disorder, schizophrenia, and normal controls. *Bipolar Disorders*, *15*(4), 422-433. https://doi.org/10.1111/bdi.12075

Siedlecki, K. L., Honig, L. S., & Stern, Y. (2008). Exploring the structure of a neuropsychological battery across healthy elders and those with questionable dementia and Alzheimer's disease. *Neuropsychology*, *22*(3), 400-411. https://doi.org/10.1037/0894-4105.22.3.400

Štěpánková, H., Bezdíček, O., Nikolai, T., Horáková, K., Lukavský, J., & Kopeček, M. (2015). National normative study of cognitive determinants of healthy ageing - status report. *E-Psychologie, 9*(1), 43-64.

Tuokko, H. A., Chou, P. H. B., Bowden, S. C., Simard, M., Ska, B., & Crossley, M. (2009). Partial measurement equivalence of French and English versions of the Canadian Study of Health and Aging neuropsychological battery. *Journal of the International Neuropsychological Society*, *15*(3), 416-425. https://doi.org/10.1017/S1355617709090602

Vazzana, R., Bandinelli, S., Lauretani, F., Volpato, S., Lauretani, F., Di Iorio, A., Abate, M., Corsi, A. M., Milaneschi, Y., Guralnik, J. M., & Ferrucci, L. (2010). Trail Making Test predicts physical impairment and mortality in older persons. *Journal of the American Geriatrics Society*, *58*(4), 719-723. https://doi.org/10.1111/j.1532-5415.2010.02780.x

Věchetová, G., Jarošová, Z., Orlíková, H., Bolceková, E., & Preiss, M. (2018). Assessment of cognitive functions using short repeatable neuropsychological batteries. *Česká a slovenská neurologie a neurochirurgie*, *81/114*(1), 29-36. https://doi.org/10.14735/amcsnn201829

Wechsler, D. (2010). *Wechslerova inteligenční škála pro dospělé – WAIS-III*. Hogrefe – Testcentrum.

Wechsler, D. (2011). *Wechslerova zkrácená paměťová škála – WMS-IIIa*. Hogrefe – Testcentrum.

Weintraub, S., Besser, L., Dodge, H. H., Teylan, M., Ferris, S., Goldstein, F. C., Giordani, B., Kramer, J., Loewenstein, D., Marson, D., Mungas, D., Salmon, D., Welsh-Bohmer, K., Zhou, X.-H., Shirk, S. D., Atri, A., Kukull, W. A., Phelps, C., & Morris, J. C. (2018). Version 3 of the Alzheimer disease centers' neuropsychological test battery in the uniform data set (UDS): *Alzheimer Disease & Associated Disorders*, *32*(1), 10-17. https://doi.org/10.1097/WAD.0000000000000223

Weintraub, S., Salmon, D., Mercaldo, N., Ferris, S., Graff-Radford, N. R., Chui, H., Cummings, J., DeCarli, C., Foster, N. L., Galasko, D., Peskind, E., Dietrich, W., Beekly, D. L., Kukull, W. A., & Morris, J. C. (2009). The Alzheimer's disease centers' uniform data set (UDS): The neuropsychologic test battery. *Alzheimer Disease & Associated Disorders*, *23*(2), 91-101. https://doi.org/10.1097/WAD.0b013e318191c7dd

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*(1), 10-18. https://doi.org/10.1111/j.1750-8606.2009.00110.x

Zaninotto, P., Batty, G. D., Allerhand, M., & Deary, I. J. (2018). Cognitive function trajectories and their determinants in older people: 8 years of follow-up in the English Longitudinal Study of Ageing. *Journal of Epidemiology and Community Health*, *72*(8), 685-694. https://doi.org/10.1136/jech-2017-210116

Zemanová, N., Bezdíček, O., Michalec, J., Nikolai, T., Roth, J., Jech, R., & Růžička, E. (2016). Validační studie české verze Bostonského testu pojmenování. *Česká a slovenská neurologie a neurochirurgie*, *79/112*(3), 307-316. https://doi.org/10.14735/amcsnn2016307

SOUHRN

Konkurenční faktorové modely kognice zdravých seniorů: podpora pro identifikaci SuperAgerů

Cíle. Neuropsychologické testy sledují v různé míře řadu kognitivních funkcí. Faktorové struktury různých neuropsychologických baterií a jejich analýzy tak vykazují jak podobnosti, tak rozdíly. Studie zkoumá českou komplexní neuropsychologickou baterii pro SuperAgery (starší osoby s vynikajícími kognitivními schopnostmi) z průřezového a longitudinálního hlediska s ohledem na její faktorovou strukturu a její stabilitu v čase.

Soubor. Soubor tvořilo 361 zdravých starších osob (ve věku 60–94 let), které byly v letech 2012 a 2015 hodnoceny pomocí baterie kognitivních testů.

Statistické analýzy. Data byla analyzována konfirmační faktorovou analýzou a analýzou invariance v čase s využitím několika konkurenčních modelů kognice založených na teoriích vycházejících z předchozích studií a sestávajících z 1-5 faktorů.

Výsledky. Výsledky ukazují, že nejlepší model se skládá ze čtyř faktorů: verbální paměť, pozornost/pracovní paměť, exekutivní funkce a jazyk. Tento model kognice byl také nejstabilnější. Model odráží kognitivní funkce zdravých seniorů a zdůrazňuje potřebu identifikovat SuperAgery na základě výkonu ve více kognitivních oblastech. Autoři navrhují, aby tyto čtyři domény byly brány v úvahu při identifikaci SuperAgerů a aby porovnávání konkurenčních modelů bylo standardním postupem v budoucích studiích.

Omezení. Ve studii nebyly zastoupeny faktory vizuálně-prostorové funkce nebo neverbální paměti s příslušnými testy. Vzorek tvořili zdraví starší dospělí.