Original Research

SAGE Open
July-September 2020: 1–10
© The Author(s) 2020
DOI: 10.1177/2158244020948846
journals.sagepub.com/home/sgo
SAGE

Latent Structure of Executive Functioning/Learning Tasks in the CogState Computerized Battery

Elisabeth Nordenswan[1,2], Eeva-Leena Kataja[1,3], Kirby Deater-Deckard[4], Riikka Korja[1,3], Mira Karrasch[2], Matti Laine[2], Linnea Karlsson[1,5], and Hasse Karlsson[1,6]

Abstract

This study tested whether executive functioning (EF)/learning tasks from the CogState computerized test battery show a unitary latent structure. This information is important for the construction of composite measures on these tasks for applied research purposes. Based on earlier factor analytic research, we identified five CogState tasks that have been labeled as EF/learning tasks and examined their intercorrelations in a new sample of Finnish birth cohort mothers (N = 233). Using confirmatory factor analyses, we compared two single-factor EF/learning models. The first model included the recommended summative scores for each task. The second model exchanged summative scores for first test round results for the three tasks providing these data, as initial task performance is expected to load more heavily on EF. A single-factor solution provided a good fit for the present five EF/learning tasks. The second model, which was hypothesized to tap more onto EF, had slightly better fit indices, $\chi^2(5) = 1.37$, $p = .93$, standardized root mean square residual (SRMR) = .02, root mean square error of approximation (RMSEA) = .00, 90% CI = [.00–.03], comparative fit index (CFI) = 1.00, and more even factor loadings (.30–.56) than the first model, $\chi^2(5) = 4.56$, $p = .47$, SRMR = .03, RMSEA = .00, 90% CI = [.00–.09], CFI = 1.00, factor loadings (.20–.74), which was hypothesized to tap more onto learning. We conclude that the present CogState sum scores can be used for studying EF/learning in healthy adult samples, but call for further research to validate these sum scores against other EF tests.

Keywords

executive functions, learning, factor analysis, CogState, neuropsychology

Introduction

Executive functions (EFs) encompass working memory, set-shifting, and inhibition—abilities that are central for human functioning by coordinating and controlling cognitive processes during complex tasks. As EFs include separable abilities that differentially contribute to executive functioning (EF) task performance, it is preferable to base EF assessment on multiple tasks (Friedman & Miyake, 2017). Composite measures are superior to single task scores, as the latter show more variability due to measurement error (Cuevas et al., 2014) and conflate task-specific and domain-general variance. Thus, there is a need for feasible EF composites based on several tasks, which can be utilized for applied purposes. In this study, we addressed this issue by exploring with confirmatory factor analysis (CFA) whether such a composite can be formed from CogState tasks that previous factor analytic studies have linked to EFs and to the related domain of learning (Chou et al., 2015; Lees et al., 2015; Yoshida et al., 2011; Zhong et al., 2013).

CogState is a computerized neuropsychological test battery designed for repeated assessments with minimal practice effects (www.cogstate.com; Pietrzak et al., 2008). Its customizable range of tasks represents computerized adaptations of standard neuropsychological tests (Pietrzak,

[1]The FinnBrain Birth Cohort Study, Turku Brain and Mind Center, Department of Clinical Medicine, University of Turku, Turku, Finland
[2]Department of Psychology, Åbo Akademi University, Turku, Finland
[3]Department of Psychology, University of Turku, Turku, Finland
[4]Psychological and Brain Sciences, University of Massachusetts Amherst, Amherst, USA
[5]University of Turku and Turku University Hospital, Department of Child Psychiatry, Turku, Finland
[6]University of Turku and Turku University Hospital, Department of Psychiatry, Turku, Finland

Corresponding Author:
Elisabeth Nordenswan, The FinnBrain Birth Cohort Study, Turku Brain and Mind Center, Department of Clinical Medicine, University of Turku, Teutori, Lemminkäisenkatu 3A, 20520 Turku, Finland.
Email: elisabeth.nordenswan@abo.fi

Olver, et al., 2009), and these tasks have been designed to measure EFs, learning and memory, visuomotor functioning, processing speed, attention, and social cognition (Maruff et al., 2009). Stimuli are mostly nonverbal and include pictures, mazes, and playing cards, minimizing cultural bias (Zhong et al., 2013). A benefit with CogState is that for some tasks, different outcome measures can be utilized depending on which neuropsychological ability is being measured (see, for example, Harel et al., 2014; Pietrzak et al., 2008). Like other computerized tests, CogState offers increased reliability and standardization of stimulus presentation, more accurate reaction time measurements, and a greater ease of administration in comparison to examiner administered pen and paper tests, making it a good alternative for time-efficient testing of large groups.

Previous factor analytic studies on CogState's latent structure have utilized exploratory factor analysis (EFA) (Chou et al., 2015; Lees et al., 2015; Yoshida et al., 2011; Zhong et al., 2013). In these studies, four CogSate tasks (the International Shopping List Test [ISL]; the Continuous Paired Associate Learning Test [CPAL]; the Groton Maze Learning Test [GML]; and the Two-Back Test [TWOB]) have loaded on a common factor, variably labeled as "EF/Learning," "Memory/Reasoning," or "Memory." That these tasks and their common factor would tap both EFs and learning appears reasonable, as they are all cognitively taxing, require explicit memory and learning, and represent novel tasks except for ISL where, in turn, executively controlled implementation of learning strategies would be beneficial.

However, utilizing a composite measure including these tasks based only on the premise that they group together in EFAs is problematic. As described by Brown (2015), EFA is an exploratory, data-driven approach, in which the factor structure is not guided by theoretical grounds. Preliminary EFAs are therefore often followed-up by more detailed CFAs, which have a greater emphasis on hypothesis testing and theory, requiring that the to-be-tested model is based on past theory and evidence (Brown, 2015). The studies that have found ISL, CPAL, GML, and TWOB to group together (Chou et al., 2015; Lees et al., 2015; Yoshida et al., 2011; Zhong et al., 2013) have also included variance from other CogState tasks in the EFAs, resulting in differing factor structures and creating uncertainty regarding the robustness of the test battery's factor structure. In addition, these EFAs do not offer a more detailed insight into the tasks' common variance besides the researchers' assumption that it is EF/learning related. Although this assumption seems reasonable considering the nature of the tasks, a more detailed examination through CFA is called for before utilizing EF/learning-related CogState tasks in a composite score.

In this study, we examined the common variance of CogState EF/learning tasks through CFA using a single-factor EF/learning model. Of the tasks that previous EFA studies have linked to EF/learning (Chou et al., 2015; Lees et al., 2015; Yoshida et al., 2011; Zhong et al., 2013), we selected the ones which based on previous research are most likely to tap onto this construct, that is, ISL, CPAL, GML, and TWOB (see the section *Method: Measures* for more details on the tasks' psychometric properties). In addition, we included a task in our CFA that was not utilized in the previous EFAs, namely, the Set-Shifting Test (SETS) that is assumed to measure the capacity for flexible shifting between tasks, a central component of EF (Miyake et al., 2000). Acknowledging the unity and diversity of EFs (Friedman & Miyake, 2017) and their important role in learning (e.g., Duff et al., 2005; Tremont et al., 2000), we hypothesized that the tasks would covary sufficiently to load on a common factor, while expecting modest to moderate standardized loadings.

For more nuanced knowledge about the shared variance of these five EF/learning tasks, we ran two separate CFAs including the same tasks, but with different task outcome variables. In terms of executive load, the novelty of a task is relevant; repeated encounters can reduce a task's effectiveness in capturing EFs (Miyake et al., 2000). Hence, for tasks including repeated trials, data from initial trials should tap more onto EF capacity than the sum score of all trials, while the sum score should tap more onto learning ability as repetition enables a learning process. The first CFA model, hypothesized to tap more onto learning, included the sum score of all trials for the tasks with repeated trials (i.e., ISL, CPAL, and GML). In contrast, the second CFA model, hypothesized to tap more onto EFs, included only the initial trials for these tasks. We furthermore explored the nature of our single-factor EF/learning models by probing correlations between sum scores based on the models and the participants' age, education, and verbal intelligence. The relationship between age and EF task performance has been described as an inverted U-shaped curve across the lifespan (Zelazo et al., 2004). In turn, strong linear associations have been found between EF and verbal intelligence (Friedman et al., 2006). Furthermore, intelligence (encompassing EFs) has strong linear associations with educational attainment (Deary & Johnson, 2010). Thus, education and verbal intelligence were expected to show positive linear correlations with the composites. As the age range of the present sample (i.e., mothers of young children) should hit the top region of the inverted U-shaped age/EF function, we anticipated moderate quadratic rather than linear correlations between the composites and age.

To the best of our knowledge, neither CogState EF/learning tasks nor the inclusion of the initial test rounds from the ISL, CPAL, and GML in a model have previously been examined through CFA. Therefore, this study offers novel information that can help to guide clinicians and researchers who want to avoid error variance related to measurements with single tasks (Cuevas et al., 2014) using a composite CogState EF/learning measure instead.

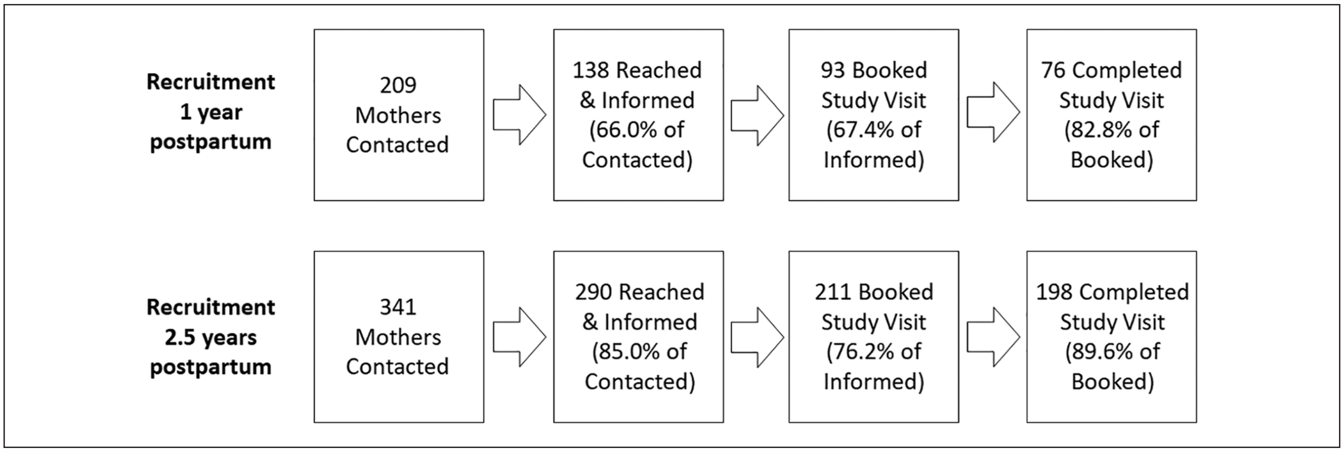

Figure 1. Participant recruitment process.
Note. Participants were eligible for this study if they spoke Finnish and were part of the FinnBrain substudy exploring the role of maternal cognition in early parenthood. Recruitment and testing were carried out between 2016 and 2018. The mothers who were contacted but did not complete a study visit were not reached, chose not to participate, or canceled a booked study visit. Forty-one participants completed assessments at both 1 and 2.5 years postpartum. For them, the first assessment was used; hence, $N = 233$ participants. Of the participants reached 1 year postpartum, those who completed a study visit did not differ significantly from those who did not in terms of age as determined by an independent samples *t*-test, $t(136) = -1.58$, $p = .12$, or education level as determined by a one-way analysis of variance, $F(1, 130) = 3.39$, $p = .07$. Of the participants reached 2.5 years postpartum, those who completed a study visit did not differ significantly from those who did not in terms of age, $t(288) = -1.05$, $p = .29$, or education level, $F(1, 285) = 1.25$, $p = .26$.

Method

Participants

Participants for this study ($N = 233$ women) were from the FinnBrain Birth Cohort ([Karlsson et al., 2018; www.finnbrain.fi]) and in a substudy exploring the role of maternal cognition in early parenthood. The FinnBrain Birth Cohort was recruited in Southwest Finland between 2011 and 2015, at maternal welfare clinics. Sufficient knowledge of Finnish or Swedish and a normal pregnancy ultrasound screening result were inclusion criteria. The coverage of women informed about the Birth Cohort study at gestational Week 12 is close to 100%. Of those informed about the study, 3,808 (66%) participated (Karlsson et al., 2018). From the FinnBrain Birth Cohort, participants were randomly selected from 2012 to 2013 and recruited to take part in a substudy exploring the role of maternal cognition in early parenthood. Exclusion criteria were self-reported neurologic or psychiatric illness and self-reported insufficient Finnish language skills. The mothers attended the substudy's first study visit during pregnancy and were invited back for follow-up study visits 1 and 2.5 years postpartum. At the 2.5-year study visit, the sample was enriched with mothers whose children had taken part in a substudy exploring the early development of self-regulation. Because the CogState test battery utilized at the follow-up visits was more extensive than the test battery utilized during pregnancy, data from the 1 year ($n = 76$) and 2.5 years postpartum ($n = 157$) measurements were selected for this study. Participant recruitment for the sample included in this study is described in Figure 1.

The participants' mean age was 33.0 years ($SD = 4.7$ years, range = 21.8–46.3 years). Almost half (44.8%) had a university-level education, 33.5% had a polytechnics education, while 21.7% had a high school/vocational education (<12 years).

Measures

The CogState test battery. CogState tasks are computerized adaptions of standard neuropsychological tests (Pietrzak, Olver, et al., 2009). Participants completed a 12-task CogState test battery, from which five tasks that measure EFs and verbal/visuospatial learning were selected for this study.

The International Shopping List Test (ISL) is a verbal list learning task, a neuropsychological method often used to assess verbal memory. The participant is instructed to remember a shopping list of 12 items that is read out loud and is then asked to recall the items. The same procedure, with the same list, is repeated three times. For Model A, we used the CogState recommended total number of correct responses from all three trials as outcome variable. To capture a greater degree of EF-related variance, the number of correct responses from the first round was chosen for Model B. The ISL has been developed to suit individuals from different linguistic/cultural backgrounds, possesses little practice effects due to the use of multiple parallel lists, has a high test–retest reliability, and has been found to be sensitive to verbal memory impairment in groups of individuals at varying stages of Alzheimer's disease (Lim et al., 2009; Lim, Harrington, et al., 2012; Lim, Pietrzak, et al., 2012; Thompson et al., 2011). The ISL's convergent validity is supported through studies showing strong correlations with the Hopkins Verbal learning Test–Revised (Hammers et al., 2013; Pietrzak, Olver, et al., 2009). Furthermore, the ISL has been found to not correlate with the verbal tasks from the Wechsler Adult Intelligence Scale (Fourth Edition) in an adult, general population sample, indicating its independence from crystallized, acquired language abilities (Kataja et al., 2017). Interestingly, Tremont and colleagues (2000) found that performance on a similar word list learning task, the California Verbal Learning Test, was sensitive to the degree of executive dysfunction.

The Continuous Paired Associate Learning Test (CPAL) is based on the visual paired associate learning paradigm, which measures the ability to encode sets of associations between spatial locations and simple patterns into memory, so that later exposure to one aspect of that same information stimulates recall of the other. First, the participant is taught the hidden locations of two differently shaped and colored figures beneath neutrally colored circles on the screen. Next, the participant is taught where eight other figures are located on the screen. As the figures are shown at the center of the screen one at a time, the participant is to find that figure by clicking on the peripheral circle under which it is hidden. During the test phase, the same eight figures are hidden under circles. Two additional, empty circles are also present. The differently shaped and colored figures are shown one by one at the center of the screen, and the participant is to select the circle behind which the figure is hidden. Incorrect responses result in an error sound, and the correct response is required to proceed. During the six test rounds, the figures are presented in differing orders. The CogState recommended total number of errors across all rounds was Model A's outcome variable, while the number of errors from the first test round (i.e., the round after the first learning trial) was chosen for Model B to tap more onto EF. The CPAL is considered to measure both EF and learning (Harel et al., 2011, 2014; O'Donnell et al., 2011). Moderate correlations between the CPAL/the Rey Auditory Verbal Learning Test, the Brief Visuospatial Memory Test–Revised, the Montreal Cognitive Assessment, and the Japanese language version of the Brief Assessment of Cognition in Schizophrenia have been reported (Chou et al., 2015; Yoshida et al., 2011; Racine et al., 2016). Furthermore, Kataja et al. (2017) reported moderate correlations between the CPAL and the WAIS-IV Matrix Reasoning task.

The Groton Maze Learning Test (GML) is based on earlier hidden maze tasks. A 28-step pathway is hidden among 100 possible locations in a 10×10 grid of tiles on the screen. After first learning the task rules in a smaller practice grid, the participant guesses the pathway from the top left corner to the bottom right corner by clicking on one tile at a time. Feedback is provided in the form of a green checkmark or a red cross on the tile after each move. After an incorrect move, the participant must click on the last correct tile and then make a different choice. The task is repeated 5 times, with the same pathway. For Model A, the CogState recommended total number of errors from all rounds was chosen as outcome variable. To tap more onto EF, the number of errors from the first test round (i.e., the round after the first learning trial) was chosen for Model B. The GML requires the ability to create, update, and efficiently access a visuospatial map of a well-defined stimulus field, measuring EF and visuospatial learning/memory (Pietrzak et al., 2007). Its convergent validity has been explored in healthy adult samples through comparisons with traditional neuropsychological tests measuring working memory, route selection, and planning/problem-solving, like the Paced Auditory Serial Addition Test, The Tower of Toronto, The Tower of Hanoi, the Two-Back Test, the WAIS-IV Working Memory Index, the Zoo Trip Test, The Mazes Test, the Rey–Osterrieth test, and the Benton test. The relationships between GML and a virtual environment navigation learning task have also been explored. Associations have varied from moderate to strong (Kataja et al., 2017; Pietrzak et al., 2007; Pietrzak, Maruff, & Snyder, 2009; Tippett et al., 2009).

The Two-Back Test (TWOB) is based on the *n*-back paradigm, a working memory task which has been widely used in neuroimaging studies and in aging research. The participant is to decide whether the playing card presented at the center of the screen is identical to the one presented two cards back. Correct/incorrect responses result in different sounds and card movements. Different cards are shown during the practice phase and the actual task. The task terminates after 32 correct responses. The CogState recommended arcsine transformation of the square root of the proportion of correct responses was chosen as outcome variable for both Models A and B. In young adults (20–40 years), *n*-back performance has been found to require EF, more specifically interference control, task switching, and updating (Gajewski et al., 2018). TWOB has been reported to correlate moderately with Stroop color–word interference, Trail Making Test Parts A & B, and WAIS-R Digit Symbol (Racine et al., 2016), moderately to strongly with the Immediate Memory, Delayed Memory, and Visuospatial/Constructional measures of the Chinese version of the Repeatable Battery for the Assessment of Neuropsychological Status (Zhong et al., 2013), and strongly with the Letter Number Span and Spatial Span from the Wechsler Memory Scale III (Pietrzak, Olver, et al., 2009). As part of CogState's Schizophrenia Battery, the TWOB has been found to be sensitive to schizophrenia-related cognitive impairment (Pietrzak, Olver, et al., 2009; Zhong et al., 2013).

The Set-Shifting Test (SETS) is similar to the Wisconsin Card Sorting Test, a neuropsychological test frequently used to measure EFs, including updating of information in working memory, suppressing irrelevant information and inhibiting prepotent responses, shifting, planning, and monitoring and controlling behavior (Rhodes, 2004). A playing card is shown at the center of the screen. The participant must guess whether the card contains a target stimulus (a color or number). Sounds indicate whether the response was correct, and the next stimulus is not displayed until a correct response has been made. The participant is in this way taught the correct card dimension. The card dimension changes after a while, and the new rule must be learnt to proceed. The task terminates after 120 correct responses. For Model A, the CogState recommended total number of errors was selected as outcome variable, while the arcsine transformation of the square root of the proportion of correct responses was chosen for Model B, as it was better distributed than the number of errors. The two measures correlate perfectly ($r_s = -1.00$, $p = .00$), capturing the same variance. In a sample of first-episode psychosis patients, the participants' SETS results were similar to the results of the Trail Making Test Part B (Benoit et al., 2015). The SETS has been used to study EFs, for example, in relation to parental reflective functioning (Rutherford et al., 2018) and as a cognitive correlate when examining the impact of long-acting injection versus oral risperidone on white matter volume (Bartzokis et al., 2011).

The Wechsler Adults Intelligence Scale–Fourth Edition, Verbal Comprehension Index (WAIS-IV VCI). The WAIS-IV (Wechsler, 2008, 2012) is a widely used intelligence tests for adults. In this study, we employed the Verbal Comprehension Index (VCI), which is derived from the verbal subtests Similarities, Information, and Vocabulary.

Procedure

The Joint Ethics Committee of Turku University Hospital and University of Turku gave ethical approval for this study. The participants gave written informed consent before participation. The testing was conducted by graduate students in quiet examination rooms, and the session took approximately 90 min. The CogState tasks were administered on a laptop computer, while the WAIS-IV VCI tasks were administered verbally. Feedback of CogState results was offered.

Data Analysis

The CFA was performed with MPlus (version 8), and all other analyses with SPSS (version 25). Mean values with standard deviations for the CogState tasks ($N = 233$) and the WAIS-IV VCI ($N = 216$) were calculated, along with the CogState completion pass rate and integrity pass rate. The WAIS-IV subtest scaled scores and the VCI scores were calculated using Finnish norms (WAIS-IV, 2012). The possible confounding effect of measurement time was examined by comparing the CogState scores (Mann–Whitney U test) and the WAIS-IV VCI scores (independent samples t-test) for the two groups tested 1 year/2.5 years postpartum.

Two EF/learning single-factor models containing the same five CogState EF/learning tasks were compared. Model A included the CogState-recommended primary outcome variables, while partly different variables hypothesized to tap more onto EF were chosen for Model B. See the task descriptions in the "Measures" section for more details about the outcome variables included in Models A and B. Both models' GML and CPAL variables, and Model A's SETS variable, were transformed so that a higher value equaled a better result for all variables.

In accordance with Brown (2015), the fit of the two single-factor models was assessed with the following fit indices and cut-off criteria: χ^2 (along with significance level and degrees of freedom), standardized root mean square residual (SRMR, values close to .08 or below), comparative fit index (CFI, values close to .95 or greater), and root mean square error of approximation (RMSEA, values close to .06 or below).

In correspondence with Models A and B, Sum Score A and Sum Score B were created by combining the standardized Z-scores for the task variables in question. To explore the sum scores' relationship to EF/learning-related variables, correlational analyses were performed between the participants' age, education level, WAIS-IV VCI, the sum scores, and the single task scores. Correlations between participant age, education level, and WAIS-VCI were also computed. Sum score–age scatterplots were created, and both the linear and the quadratic slopes were plotted.

Results

The mean values and standard deviations of the CogState tasks, the WAIS-IV tasks, and VCI are presented in Table 1. The participant's verbal intelligence corresponded with the general Finnish population. The CogState task completion rate was 100% for all tasks except for SETS. Two participants' SETS results were excluded as they were incomplete. The CogState tasks' integrity pass rate was 100% for all tasks except for TWOB. Seven participants' TWOB results were excluded due to an insufficient pass rate.

The CogState task results for mothers tested 1 versus 2.5 years postpartum were similar (U-tests, $p = .21$–$.95$), except for SETS in both Model A ($U = 7{,}185.00$, $p = .01$) and Model B ($U = 4{,}595.00$, $p = .01$). The mothers tested with SETS 1 year postpartum made fewer errors ($M = 15.8$, $SD = 9.7$) than the mothers tested 2.5 years postpartum ($M = 21.0$, $SD = 13.9$). The two groups' WAIS-IV VCI results did not differ significantly ($t = -.52$, $df = 214$, $p = .60$, two-tailed).

The loadings and error terms of the two single-factor EF/learning models are presented in Figure 2. The fit indices for both Model A, $\chi^2(5) = 4.56$, $p = .47$, SRMR $= .03$, RMSEA $= .00$, 90% CI $= [.00$–$.09]$, CFI $= 1.00$, and Model B, $\chi^2(5) = 1.37$, $p = .93$, SRMR $= .02$, RMSEA $= .00$, 90% CI $= [.00$–$.03]$, CFI $= 1.00$, demonstrate that they provide a good fit with the data. However, Model B had slightly better model fit indices, and more even factor loadings, indicating that it better captures the common variance of the five tasks.

The correlational analyses are presented in Table 2. Sum Score A and Sum Score B overall correlated fairly similarly with participant age, education level, and verbal intelligence. As expected, education level and WAIS-IV VCI correlated positively with the sum scores, and with some of the single tasks. Furthermore, when the quadratic curve was plotted in

Table 1. Mean Values and Standard Deviations of CogState Tasks/WAIS-IV Tasks and VCI.

CogState tasks	*M* (*SD*), *N* = 233
ISL, number of correct responses, all test rounds[a]	29.34 (3.31)
ISL, number of correct responses, first test round[a]	7.93 (1.56)
CPAL, amount of errors, all test rounds[b]	40.52 (34.43)
CPAL, amount of errors, first test round[b]	12.64 (8.83)
GML, amount of errors, all test rounds[b]	38.80 (11.22)
GML, amount of errors, first test round[b]	8.46 (3.70)
TWOB, arcsine transformation of the proportion of correct responses[a]	1.28 (0.18)
SETS, amount of errors[b]	19.30 (12.90)
SETS, arcsine square root of the proportion of correct responses[a]	1.21 (0.10)
WAIS-IV	*M* (*SD*), *N* = 216
Similarities	10.42 (3.02)
Vocabulary	10.26 (3.22)
Information	9.84 (3.30)
VCI[a]	100.99 (15.88)

Note. WAIS-IV = the Wechsler Adults Intelligence Scale–Fourth Edition; VCI = Verbal Comprehension Index; ISL = International Shopping List Test; CPAL = Continuous Paired Associate Learning Test; GML = Groton Maze Learning Test; TWOB = Two-Back Test; SETS = Set-Shifting Test.
[a]Highercore = better performance. [b]Lower score = better performance.

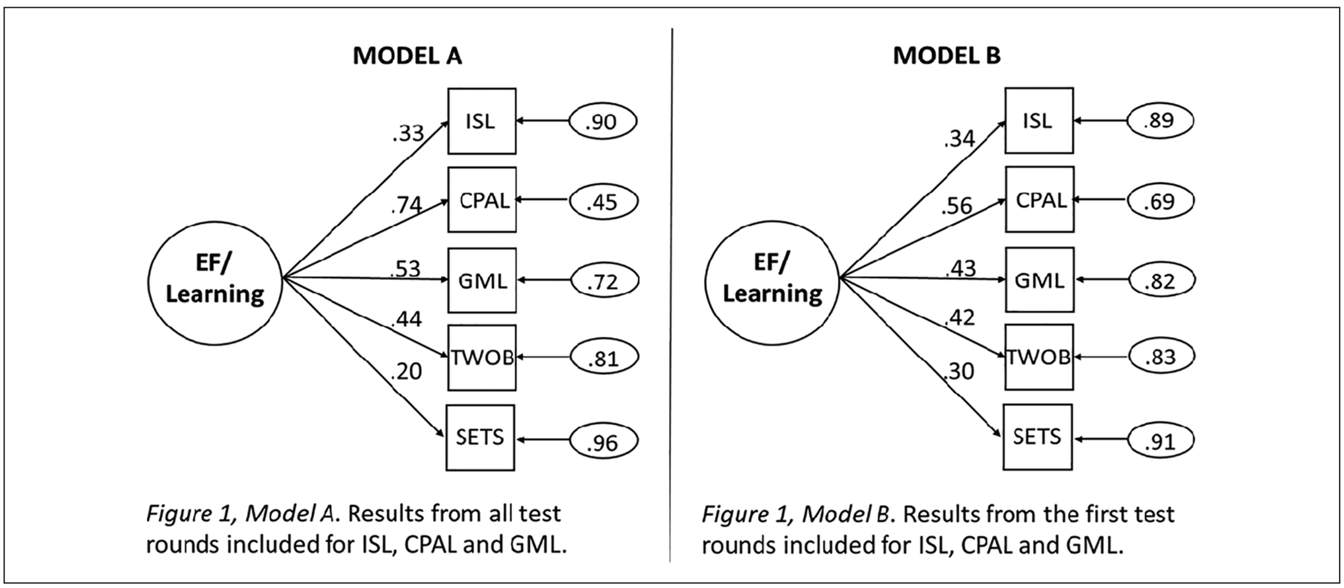

Figure 2. EF/learning factor Models A and B, factor loading, and error terms.
Note. EF = executive functioning; ISL = International Shopping List Test; CPAL = Continuous Paired Associate Learning Test; GML = Groton Maze Learning Test; TWOB = Two-Back Test; SETS = Set-Shifting Test.

the sum score–age scatterplots (see Figure 3), the expected inverted U-shaped curve appeared. For both sum scores, the model with the quadratic (and linear) term fitted significantly better than the model with only the linear term (Sum Score A: $p = .01$, R^2 change = .03; Sum Score B: $p = .01$, R^2 change = .03).

Discussion

This study examined the latent structure of five EF/learning tasks from CogState, a computerized neuropsychological test battery (www.cogstate.com; Pietrzak et al., 2008). We examined whether these tasks could be combined into a single composite, as previous exploratory factor analyses on CogState tasks have suggested (Chou et al., 2015; Lees et al., 2015; Yoshida et al., 2011; Zhong et al., 2013). This study was further motivated by previous research indicating close links between EF and learning measures (Duff et al., 2005; Tremont et al., 2000). We tested this with a sample of healthy birth cohort mothers using a CFA. Our novel approach compared two sets of variables from the five CogState EF/learning tasks: one including the recommended summative scores

Table 2. Correlations Between Sum Scores/Single Tasks and Age/Education Level/WAIS-IV VCI.

Correlational variables	Age	Education level	WAIS-IV VCI
Sum Score A	−.16**	.20**	.21**
Sum Score A's ISL	−.03	.19**	.23**
Sum Score A's CPAL	−.22**	.08	.08
Sum Score A's GML	−.17**	.05	.00
Sum Score A's TWOB	−.06	.10	.14*
Sum Score A's SETS	−.05	.11	.13*
Sum Score B	−.12*	.20**	.25**
Sum Score B's ISL	−.05	.13*	.22**
Sum Score B's CPAL	−.19**	.05	.09
Sum Score B's GML	−.04	.14*	.12*
Sum Score B's TWOB	−.06	.10	.14*
Sum Score B's SETS	−.05	.11	.13*
Age	1		
Education level	.30**	1	
WAIS-IV VCI	.28**	.52**	1

Note. For correlations involving the variables education level, TWOB, or SETS, Spearman's rho was used; for other correlations, Pearson's *r* was used. For the sum scores, the single CogState tasks, and WAIS-IV VCI, a higher value means a better result. WAIS-IV = the Wechsler Adults Intelligence Scale–Fourth Edition; VCI = Verbal Comprehension Index; ISL = International Shopping List Test, CPAL = Continuous Paired Associate Learning Test, GML = Groton Maze Learning Test, TWOB = Two-Back Test, SETS = Set-Shifting Test.
*Correlations significant at the .05 level. **Correlations significant at the .01 level (all one-tailed).

and the other including first test rounds for the three tasks for which these data were available (ISL, CPAL, GML). The latter set of variables was hypothesized to reflect more EF load. In addition, we probed correlations between the two EF/learning sum scores and the participants' age, education, and verbal intelligence. Taken together, the results indicated that (a) the five EF/learning tasks (ISL, CPAL, GML, TWOB, and SETS) are combinable into an EF/learning sum score, (b) the choice of variables (whole test performance vs. first round) had an effect on the single-factor model properties, and (c) the composite scores correlated with selected background variables in the expected fashion.

Both models' fit indices showed a good fit with the data, supporting the construction of an EF/learning sum score that would allow for more reliable and versatile assessment compared with single tasks. In accordance with the unity and diversity of EFs (Friedman & Miyake, 2017) and their role in learning (Tremont et al., 2000), most factor loadings were either within the expected range of .20 to .40, or very close to it. Together with previous knowledge about the psychometric properties of the CogState tasks (see the task descriptions in the "Measures" section), these CFA results indicate that combined with the SETS, the CogState tasks which in previous exploratory factor analyses have grouped together (Chou et al., 2015; Lees et al., 2015; Yoshida et al., 2011; Zhong et al., 2013) capture a considerable amount of common variance. As stated in the *Introduction*, it seems reasonable that these tasks would tap onto both EF and learning, as they are cognitively taxing, require explicit memory and learning, and/or represent novel tasks.

A comparison between Models A and B shows that the choice of variables for ISL, CPAL, and GML (whole test performance vs. first round) affects the properties of the single-factor model. Overall, Model B's fit indices were slightly better, and its factor loadings were more even, suggesting that Model B somewhat better captures common variance among the included tasks than Model A. In Model A, CPAL (.74) and GML (.53) have especially high factor loadings, while a particularly low factor loading was found for the set-shifting task (SETS; .20) that measures a central component of EF. CPAL and GML both have a large visuospatial learning component, and the repeated ISL/CPAL/GML trials enable learning through repetition. Thus, Model A is likely to primarily measure visuospatial learning. In comparison, in Model B, CPAL (.56) and GML (.43) have lower factor loadings, while SETS (.30) has a higher factor loading. Considered together with the fact that Model B includes the first attempts for three tasks (and novel tasks are known to tap onto EFs to a higher degree than repeated tasks, Miyake et al., 2000), these factor loadings suggest that an executive component would have more weight in Model B. To summarize, our results suggest that the selection of different outcome variables for ISL/CPAL/GML affects the way they tap different cognitive domains, which is in line with previous studies where different outcome measures from CogState tasks have been utilized (e.g., Harel et al., 2014; Pietrzak et al., 2008). Model A seems to primarily measure (visuospatial) learning, while Model B appears to have a stronger executive component.

As hypothesized, the sum score–age correlations were quite low, and the scatterplots' quadratic slopes exhibited a moderate inverted U-shaped function. This conforms to the expectation that the age range of our sample hits the top of the inverted U-shaped function previously found between

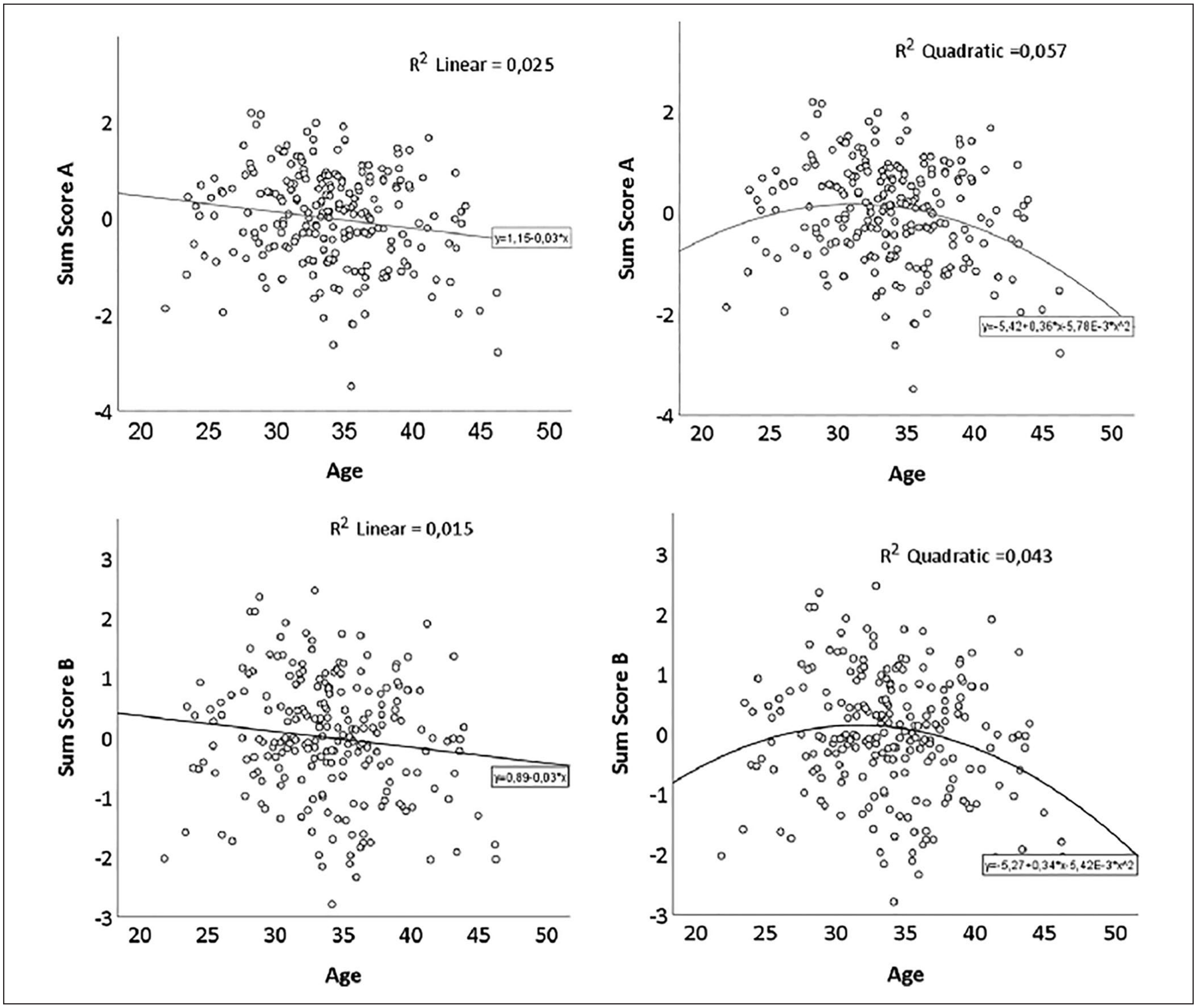

Figure 3. Linear and quadratic scatter plots between Sum Score A and age/Sum Score B and age.

age and EF (Zelazo et al., 2004). The similar correlations between the sum scores and education level/WAIS-IV VCI are logical, as intelligence and education level interact. This association is further underscored by the purely verbal task ISL showing the most consistent correlations with both education level and WAIS-IV VCI, and by the high correlation between education level and WAIS-IV VCI (r_s = .52**). Furthermore, the low correlations between the sum scores and WAIS-IV VCI demonstrate that they reflect other abilities than crystallized verbal IQ. To summarize, the correlational results support the idea that the present single factor reflects the domains of EF and learning.

The main limitation of this study is the lack of validation of the EF/learning sum scores against other neurocognitive tests known to measure EF/learning. As the study sample consisted of healthy, fairly highly educated birth cohort mothers, generalizations to other populations should be made cautiously. Further studies will be needed to explore the suitability of these sum scores for different populations and to address the clinical applicability of the sum scores. The CogState test battery has developed over time, and thus, not all currently available tasks were available when this study was initiated. Therefore, future studies focusing on EFs and CogState could include, for example, the Go No-Go Test as a measure of EF/inhibition.

In conclusion, the CogState sum scores examined in this study seem suitable as EF/learning measures in healthy adult samples, thus enabling a more comprehensive and reliable assessment compared with single tasks. The sum scores capture a satisfactory amount of the common variance of the included tasks. Sum Score A appears to be better suited for primarily measuring learning, while Sum Score B seems to better capture a shared executive component. Future studies should attempt to validate these sum scores, especially against traditional EF tests, and to explore the suitability of the sum scores for different populations.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Academy of Finland (grant numbers 253270, 134950, 286829, and 308252), the Jane and Aatos Erkko Foundation, the Signe and Ane Gyllenberg Foundation, the State Research Funding of the Turku University Hospital, the OLVI Foundation, the Oscar Öflund Foundation, the C.G. Sundell Foundation, the Frans Wilhelm and Waldemar von Frenckell Fund, the Niilo Helander Foundation, and the Victoria Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Ethics Statement

The Joint Ethics Committee of Turku University Hospital and University of Turku gave ethical approval for this study (approval number: 57/180/2011). The participants gave written informed consent before participation.

ORCID iD

Elisabeth Nordenswan https://orcid.org/0000-0002-1999-6675

References

Bartzokis, G., Lu, P. H., Amar, C. P., Raven, E. P., Detore, N. R., Altshuler, L. L., . . . Nuechterlein, K. H. (2011). Long acting injection versus oral Risperidone in first-episode schizophrenia: Differential impact on white matter myelination trajectory. *Schizophrenia Research*, *132*, 35–41. https://doi.org/10.1016/j.schres.2011.06.029

Benoit, A., Malla, A. K., Iyer, S. N., Joober, R., Bherer, L., & Lepage, M. (2015). Cognitive deficits characterization using the CogState Research Battery in first-episode psychosis patients. *Schizophrenia Research: Cognition*, *2*, 140–145. https://doi.org/10.1016/j.scog.2015.03.006

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.

Chou, C. C., Pressler, S. J., Giordani, B., & Fetzer, S. J. (2015). Validation of the Chinese version of the CogState computerised cognitive assessment battery in Taiwanese patients with heart failure. *Journal of Clinical Nursing*, *24*(21–22), 3147–3154. https://doi.org/10.1111/jocn.12919

Cuevas, K., Deater-Deckard, K., Kim-Spoon, J., Watson, A. J., Morasch, K. C., & Bell, M. A. (2014). What's mom got to do with it? Contributions of maternal executive function and caregiving to the development of executive function across early childhood. *Developmental Science*, *17*(2), 224–238. https://doi.org/10.1111/desc.12073

Deary, I. J., & Johnson, W. (2010). Intelligence and education: Causal perceptions drive analytic processes and therefore conclusions. *International Journal of Epidemiology*, *39*, 1362–1369. https://doi.org/10.1093/ije/dyq072

Duff, K., Schoenberg, M. R., Scott, J. G., & Adams, R. L. (2005). The relationship between executive functioning and verbal and visual learning and memory. *Archives of Clinical Neuropsychology*, *20*, 111–122. https://doi.org/10.1016/j.acn.2004.03.003

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179. https://doi.org/10.1111/j.1467-9280.2006.01681.x

Gajewski, P. D., Hanisch, E., Falkenstein, M., Thönes, S., & Wascher, E. (2018). What does the *n*-back task measure as we get older? Relations between working-memory measures and other cognitive functions across the lifespan. *Frontiers in Psychology*, *9*, Article 2208. https://doi.org/10.3389/fpsyg.2018.02208

Hammers, D. B., Jung, M., Pressler, S. J., Sullivan, B. J., Koelling, T., & Giordani, B. (2013). Clinical utility of auditory memory testing in a heart failure population. *Journal of Cardiovascular Nursing*, *28*(5), 444–452. https://doi.org/10.1097/JCN.0b013e318258abf3

Harel, B. T., Darby, D., Pietrzak, R. H., Ellis, K. A., Snyder, P. J., & Maruff, P. (2011). Examining the nature of impairment in visual paired associate learning in amnestic mild cognitive impairment. *Neuropsychology*, *25*(6), 752–762. https://doi.org/10.1037/a0024237

Harel, B. T., Pietrzak, R. H., Snyder, P. J., Thomas, E., Mayes, L. C., & Maruff, P. (2014). The development of associate learning in school age children. *PLOS ONE*, *9*(7), Article e101750. https://doi.org/10.1371/journal.pone.0101750

Karlsson, L., Tolvanen, M., Scheinin, N. M., Uusitupa, H. M., Korja, R., Ekholm, E., … Karlsson, H. (2018). Cohort profile: The FinnBrain birth cohort study (FinnBrain). *International Journal of Epidemiology*, *47*(1), 15–16j. https://doi.org/10.1093/ije/dyx173

Kataja, E. L., Karlsson, L., Tolvanen, M., Parsons, C., Schembri, A., Kiiski-Mäki, H., & Karlsson, H. (2017). Correlation between the CogState computerized measure and WAIS-IV among birth cohort mothers. *Archives of Clinical Neuropsychology*, *32*, 252–258. https://doi.org/10.1093/arclin/acw099

Lees, J., Applegate, E., Emsley, R., Lewis, S., Michalopoulou, P., Collier, T., . . . Drake, R. J. (2015). Calibration and cross-validation of MCCB and CogState in schizophrenia. *Psychopharmacology*, *232*, 3873–3882. https://doi.org/10.1007/s00213-015-3960-8

Lim, Y. Y., Harrington, K., Ames, D., Ellis, K. A., Lachovitzki, R., Snyder, P. J., & Maruff, P. (2012). Short term stability of verbal memory impairment in mild cognitive impairment and Alzheimer's disease measured using the International Shopping List Test. *Journal of Clinical and Experimental Neuropsychology*, *34*(8), 853–863. https://doi.org/10.1080/13803395.2012.689815

Lim, Y. Y., Pietrzak, R. H., Snyder, P. J., Darby, D., & Maruff, P. (2012). Preliminary data on the effect of culture on the assessment of Alzheimer's disease-related verbal memory impairment with the International Shopping List Test. *Archives of Clinical Neuropsychology*, *27*, 136–147. https://doi.org/10.1093/arclin/acr102

Lim, Y. Y., Prang, K. H., Cysique, L., Pietrzak, R. H., Snyder, P. J., & Maruff, P. (2009). A method for cross-cultural adaptation of a verbal memory assessment. *Behavior Research Methods*, *41*(4), 1190–1200. https://doi.org/10.3758/BRM.41.4.1190

Maruff, P., Thomas, E., Cysique, L., Brew, B., Collie, A., Snyder, P., & Pietrzak, R. H. (2009). Validity of the CogState Brief Battery: Relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. *Archives of Clinical Neuropsychology*, *24*(2), 165–178. https://doi.org/10.1093/arclin/acp010

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

O'Donnell, J., Pietrzak, R. H., Ellis, K. C., Snyder, P. J., & Maruff, P. (2011). Understanding failure of visual paired associate learning in amnestic mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology*, *33*(10), 1069–1078. https://doi.org/10.1080/13803395.2011.596821

Pietrzak, R. H., Cohen, H., & Snyder, P. J. (2007). Spatial learning efficiency and error monitoring in normal aging: An investigation using a novel hidden maze learning test. *Archives of Clinical Neuropsychology*, *22*, 235–245. https://doi.org/10.1016/j.acn.2007.01.018

Pietrzak, R. H., Maruff, P., Mayes, L. C., Roman, S. A., Sosa, J. A., & Snyder, P. J. (2008). An examination of the construct validity and factor structure of the Groton Maze Learning Test, a new measure of spatial working memory, learning efficiency, and error monitoring. *Archives of Clinical Neuropsychology*, *23*(4), 433–445. https://doi.org/10.1016/j.acn.2008.03.002

Pietrzak, R. H., Maruff, P., & Snyder, P. J. (2009). Convergent validity and effect of instruction modification on the Groton Maze Learning Test: A new measure of spatial working memory and error monitoring. *International Journal of Neuroscience*, *119*, 1137–1149. https://doi.org/10.1080/00207450902841269

Pietrzak, R. H., Olver, J., Norman, T., Piskulic, D., Maruff, P., & Snyder, P. J. (2009). A comparison of the CogState Schizophrenia Battery and the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Battery in assessing cognitive impairment in chronic schizophrenia. *Journal of Clinical and Experimental Neuropsychology*, *31*(7), 848–859. https://doi.org/10.1080/13803390802592458

Racine, A. M., Clark, L. R., Berman, S. E., Koscik, R. L., Mueller, K. D., Norton, D., . . . Johnson, S. C. (2016). Associations between performance on an abbreviated CogState Battery, other measures of cognitive function, and biomarkers in people at risk for Alzheimer's Disease. *Journal of Alzheimer's Disease*, *18*(54), 1395–1408. https://doi.org/10.3233/JAD-160528

Rhodes, M. G. (2004). Age-related differences in performance on the Wisconsin card sorting test: A meta-analytic review. *Psychology and Aging*. https://doi.org/10.1037/0882-7974.19.3.482

Rutherford, H. J. V., Byrne, S. P., Crowley, M. J., Bornstein, J., Bridgett, D. J., & Mayes, L. C. (2018). Executive functioning predicts reflective functioning in mothers. *Journal of Child and Family Studies*, *27*, 944–952. https://doi.org/10.1007/s10826-017-0928-9

Thompson, T. A. C., Wilson, P. H., Snyder, P. J., Pietrzak, R. H., Darby, D., Maruff, P., & Buschke, H. (2011). Sensitivity and test-retest reliability of the International Shopping List Test in assessing verbal learning and memory in mild Alzheimer's disease. *Archives of Clinical Neuropsychology*, *26*, 412–424. https://doi.org/10.1093/arclin/acr039

Tippett, W. J., Lee, J. H., Mraz, R., Zakzanis, K. K., Snyder, P. J., Black, S. E., . . . Eng, P. (2009). Convergent validity and sex differences in healthy elderly adults for performance on 3D virtual reality navigation learning and 2D hidden maze tasks. *Cyberpsychology and Behavior*, *12*(2), 169–174. https://doi.org/10.1089/cpb.2008.0218

Tremont, G., Halpert, S., Javorsky, D. J., & Stern, R. A. (2000). Differential impact of executive dysfunction on verbal list learning and story recall. *Clinical Neuropsychologist*, *14*(3), 295–302. https://doi.org/10.1076/1385-4046(200008)14:3;1-P; FT295

Wechsler, D. (2008). *Wechsler adult intelligence scale-IV*. San Antonio, TX: Pearson.

Wechsler, D. (2012). WAIS IV esitys- ja pisteytyskäsikirja [WAIS IV manual]. Helsinki: Hogrefe Psykologien Kustannus Oy (Psychological Corporation).

Yoshida, T., Suga, M., Arima, K., Muranaka, Y., Tanaka, T., Eguchi, S., . . . Hashimoto, K. (2011). Criterion and construct validity of the CogState Schizophrenia Battery in Japanese patients with schizophrenia. *PLOS ONE*, *6*(5), Article e20469. https://doi.org/10.1371/journal.pone.0020469

Zelazo, P. D., Craik, F. I. M., & Booth, L. (2004). Executive function across the life span. *Acta Psychologica*, *115*(2–3), 167–183. https://doi.org/10.1016/j.actpsy.2003.12.005s

Zhong, N., Jiang, H., Wu, J., Chen, H., Lin, S., Zhao, Y., . . . Zhao, M. (2013). Reliability and validity of the CogState Battery Chinese language version in schizophrenia. *PLOS ONE*, *8*(9), Article e74258. https://doi.org/10.1371/journal.pone.0074258s