Primenjena psihologija
Vol. 18, No. 1, pp. 137–172, 2025

Research Article

Traditional and Computer-Based Assessment of Executive Functions

Željka Nikolašević [1] ✉, Bojana M. Dinić [2], Milan Oljača [2], Ilija Milovanović [2], Jasmina Kodžopeljić [2], and Vojislava Bugarski Ignjatović [1]

[1] *Department of Psychology, Faculty of Medicine, University of Novi Sad, Novi Sad, Serbia*

[2] *Department of Psychology, Faculty of Philosophy, University of Novi Sad, Novi Sad, Serbia*

ABSTRACT

This study aimed to assess the psychometric properties of two newly developed computer-based tasks (i.e., Mental Shifting/Flexibility Task and Auditory-Visual Go/No-Go Task) for measuring two key domains of executive functions (EF) - inhibition and mental shifting (flexibility) - in healthy adults. Together with these tasks, traditional paper-and-pencil tests were used for assessing construct validity (Wisconsin Card Sorting Test - WCST, Trail Making Test - TMT, Verbal Fluency Tests, and Advanced Progressive Matrices- APM). The sample consisted of 468 adult twins (70.7% female, mean age 24.06 years) or 234 twin pairs. Results revealed low to moderate correlations between the reaction times and the number of errors in the computer-based tasks and traditional tests. Specifically, the Mental Shifting/Flexibility Task showed significant correlations with the TMT and the WCST. The Auditory-Visual Go/No-Go Task was significantly related to TMT and APM, suggesting shared cognitive processes linked to inhibition, cognitive flexibility, and processing speed. The computer-based tasks demonstrated moderate to good ICC reliability, especially in reaction time measures, while error rates showed poorer reliability. It was concluded that computer-based tasks are useful for measuring executive functions. However, further validation, development of standardized norms, and optimization of these tools are needed. Future research should explore how these tools can be integrated into existing cognitive assessment batteries for more accurate measurement of executive functions across diverse populations and clinical contexts.

Keywords: Computer-based assessments, Executive functions, Psychometric properties, Cognitive flexibility, Inhibition

UDK: 159.938
DOI: 10.19090/pp.v18i1.2601
Received: 31.08.2024.
Revised: 03.01.2025.
Accepted: 18.02.2025.

✉ Corresponding author email: zeljka.nikolasevic@mf.uns.ac.rs

Introduction

Executive functions (EFs) refer to a set of cognitive processes involved in the control, regulation, and management of other cognitive activities. These functions allow individuals to plan, initiate, monitor, and adjust their behavior to achieve goals (Miyake et al., 2000). Initially viewed as a unified construct, research now conceptualizes EFs as a hierarchical framework comprising several lower-order functions (e.g., Diamond, 2013). Miyake's model, one of the most influential models of executive functions, identifies three core components crucial for goal-directed behavior: 1) inhibition, the ability to suppress automatic responses in favor of more appropriate actions; 2) shifting (cognitive flexibility), the ability to switch between tasks or strategies; and 3) updating, which involves maintaining and refreshing relevant information in working memory. These components are measured as latent variables inferred from multiple tasks or indicators (Miyake et al., 2000).

Traditional methods for assessing executive functions (EFs) include structured paper-and-pencil tests, clinical interviews, and behavioral observations. Well-known tools include the Wisconsin Card Sorting Test (WCST; Heaton et al., 1993), measuring cognitive flexibility and rule-shifting; the Stroop Test (Golden, 1978), assessing selective attention and inhibition; the Trail Making Test (TMT; Reitan, 1955), measuring attention and task-switching; and the Verbal Fluency Test (Goodglass & Kaplan, 1983), assessing cognitive flexibility and lexical access. Although paper-based tests have a long history and are widely accepted, they have significant drawbacks. These tests are time-consuming for clinicians to administer and score, lack flexibility in modifying tasks or stimuli, and provide limited outcome measures (Miller & Barr, 2017; Vermeent et al., 2020). While paper-based tests are supported by normative data, newer digital cognitive tests offer more efficient and flexible alternatives (e.g., Feenstra et al., 2017; Kessels, 2019; Riordan et al., 2013).

Computer-based assessment of EFs offers several advantages: 1) standardization and objectivity – it reduces human error by standardizing test administration and minimizing variability in scoring and participant interaction; 2) efficiency – these tests are faster to administer, with automated scoring and quicker result interpretation; 3) interactivity – they can incorporate various stimuli (visual, auditory, tactile), enabling a broader EF assessment; 4) precision – real-time performance monitoring enhances accuracy in measuring reaction times and task-solving speed; 5) accessibility – these tests can be used in multiple settings, including remote ones, offering greater scalability (Kane & Kay, 1992; Schatz & Browndyke, 2002). Commonly used computer-based tests include computerized versions of the WCST (Tien et al., 1996), Stroop Test (Capovilla et al., 2005), the N-back Task (Jacola et al., 2014), the Iowa Gambling Task (Dancy & Ritter, 2017), the Tower of London Task (Tybursku et al., 2021), the Go/No-Go Task (Tybursku et al., 2021), and the Flanker Task (Sanders et al., 2018).

Additionally, computer-based testing with simple cognitive tasks offers more precise measurements of EF processes, focusing on specific components like inhibition, shifting, and working memory updating. Unlike traditional tests such as the WCST and Tower of Hanoi, which suffer from "task impurity" due to non-EF factors (e.g., visual-spatial processing, memory), computer-based tasks have clearer theoretical foundations and higher reliability (Miyake et al., 2000). However, traditional tests remain crucial in clinical settings, providing valuable diagnostic insights. Thus, computer-based tasks should complement, not replace, traditional assessments.

In addition to the advantages of computer-based tests, research has focused on validating these tasks for assessing executive functions. A meta-analysis showed a strong correlation of .91 between computerized and paper-based cognitive ability tests, with higher correlations for timed power tests compared to speeded tests (Mead & Drasgow, 1993). Studies have confirmed the equivalence of computer-based and traditional versions of the WCST in healthy adults (Tien et al., 1996; Wagner & Trentini, 2009) as well as in elderly populations (Collerton et al., 2007). Similarly, the digital version of the TMT showed strong correlations with its traditional counterpart, discriminating effectively between younger and older adults (Park & Schott, 2022). However, the digital TMT required more time in older adults, likely due to cognitive load and technology familiarity (Latendorf et al., 2021). Research has also suggested that computer familiarity can modestly affect performance (McDonald, 2002). While most studies confirm the equivalence between computer-based and traditional cognitive tests (e.g., Park & Schott, 2022), some report discrepancies, highlighting the need for deeper insight into the psychometric properties of these tests (Steinmetz et al., 2010). Computer-based tasks have demonstrated convergent validity, showing strong correlations with traditional measures of cognitive flexibility and intelligence. Tasks like the Flanker Task and Tower of London are well-correlated with the WCST, Digit Span, Stroop Test, and TMT (Miyake et al., 2000). Similarly, the N-back task, used for working memory, correlates strongly with traditional neuropsychological tests (Miyake et al., 2000). However, in older adults, N-back performance is more linked to attention and memory functions than executive control, with attentional switching and updating being the most consistent cognitive functions across age groups (Gajewski et al., 2018).

Despite their advantages, computer-based methods face challenges, particularly in terms of psychometric properties like reliability and construct validity. Technological factors such as cognitive load, screen fatigue, and limited access to technology can affect performance (Alloway & Carpenter , 2020). Further research is needed to validate these methods, especially in relation to demographic and cultural factors that may influence results. While studies by Wagner and Trentini (2009) and Latendorf et al. (2021) support the reliability and sensitivity of computer-based assessments, additional validation is still required.

The Current Study

While computer-based assessment tools have demonstrated potential as valuable complements to traditional tests, their rigorous validation is crucial as new tasks continue to emerge. This study aims to assess the psychometric characteristics of newly developed computer-based tests for executive functions in healthy adults, specifically focusing on their internal consistency and construct validity. To validate these tasks, the study compares them to traditional neuropsychological assessment tools (e.g., Wisconsin Card Sorting Test, Trail Making Test), ensuring that the computer-based tools meet the same high standards for reliability and validity. Building on prior research, we hypothesize that these computer-based tasks will show moderate to strong correlations with traditional tests assessing cognitive flexibility, inhibition, and processing speed. These hypotheses are informed by previous findings indicating robust links between computerized and paper-based assessments of executive functions (e.g., Tien et al., 1996; Park & Schott, 2021).

Furthermore, this study aims to contribute theoretically by offering a deeper understanding of the cognitive processes underlying executive functions. By examining the relationships between different executive functions—such as inhibition and cognitive flexibility—this research seeks to shed light on how these processes interact in healthy adults, thereby advancing our knowledge of cognitive control mechanisms and their measurement.

Method

Participants and Procedure

A total of 468 participants, comprising 234 twin pairs, took part in the study. The majority of participants were female (70.71%), with an average age of 24.06 years (*SD* = 7.02). Regarding zygosity, most pairs were monozygotic (65.38%). Among dizygotic twin pairs, the majority (55.56%) were same-sex pairs. Detailed characteristics of the sample can be found in Appendices A and B. The Serbian Advanced Twin Registry (STAR) has established a comprehensive framework for recruitment, testing, and data collection, as described in detail by Smederevac et al. (2019). Procedures for determining twins' zygosity are outlined in Mitrović et al. (2024). Ethical approval for the study was obtained from the Institutional Ethical Committees under the codes #02-374/15, #01-39/229/1, and #O-EO-024/2020. Participation was entirely voluntary, with all participants signing informed consent forms prior to testing. Participants who volunteered were invited to attend an in-person assessment. Each measure used in this research was administered individually. The data utilized in this study were gathered over the period from 2012 to 2024.

Measures

Computer-Based Executive Functions Tasks

Mental Shifting/Flexibility Task

This task primarily measures mental shifting (flexibility) - the ability to efficiently shift attention between tasks or mental sets (Allport & Wylie, 1999; Friedman et al., 2008; Miyake et al., 2000). This involves deactivating one set and activating another based on task demands and overcoming proactive interference when new operations are hindered by previous ones. It consists of five progressively complex blocks, each increasing in complexity according to the demands placed on the participants. Each block contains 39 trials, consisting of letters (A, G, U, ...), numbers (1, 6, 7, ...), and symbols (?, *, ;, #, ...) in various colors (there were 13 combinations of stimuli). Participants were instructed that different letters, numbers, and symbols would appear alternately on the screen in different colors. Their task was to press one of the mouse buttons according to the instructions provided before the start of each block. Participants were also instructed to perform the tasks as quickly and accurately as possible. As with the first task, participants underwent a brief practice session (10 trials) before beginning the series of five blocks: 1) in the first block, participants were instructed to press the right mouse button if a letter, number, or symbol appeared in blue, and to press the left mouse button in all other situations; 2) in the second block, participants were instructed to press the right mouse button if a letter appeared, and to press the left mouse button in all other situations; 3) in the third block, participants were instructed to press the right mouse button if a number in red appeared, and to press the left mouse button in all other situations; 4) in the fourth block, participants were instructed to press the right mouse button if an odd number appeared, and to press the left mouse button in all other situations; 5) in the fifth and final block, participants were instructed to press the right mouse button if a yellow vowel appeared (A, E, i, U), and to press the left mouse button in all other situations. Instead of I, the lowercase letter i was used so as not to interfere with the number one (1), and O was not used so as not to interfere with zero (0). Therefore, each block introduces a new instruction and task demand, requiring shifts between symbols (letters, numbers, signs) and their characteristics (color, even-odd, consonant-vowel). The following variables were used in the study: average reaction time (RT) and the number of errors in each block, with faster responses and fewer errors indicating a better shifting ability.

Auditory-Visual Go/No-Go Task

This task is a variant of the standard Go/No-Go task (Garavan et al., 1999) and assesses the inhibition function, which is the ability to suppress automatic or dominant responses when necessary. The task consists of three blocks, each containing 40 trials, with progressively increasing complexity across the blocks. The stimuli are arrows pointing left or right (← or →), with or without accompanying sound. Participants were instructed to press one of two mouse buttons according to the instructions given before the start of each block. They were also instructed to perform the tasks as quickly and accurately as possible. Before starting the series of blocks, participants underwent a practice block (10 trials). In the first block, participants had to press the left button if the arrow pointed left and the right button if the arrow pointed right. During the second block, they were instructed to press the left button if the arrow pointed left and was accompanied by sound and the right button if the arrow pointed right and was accompanied by sound. In all other situations (left and right arrows without accompanying sound), participants were not to press any button. The critical signals (silent arrows) appear in a 16:40 ratio, in a pseudo-random order. After the stimulus duration of 500 ms, the next stimulus appeared. The third block required participants to press the right button only if the arrow pointed right and was not accompanied by sound. In all other situations (left arrow with or without sound and right arrow with sound), participants had to press the left button. The following variables were used in the study: average reaction time (RT) and the number of errors in each block, with faster responses and fewer errors indicating better inhibition ability. Note that in the second block, the correct response was not to press any button, so these correct responses were not included in the average reaction time, as they all corresponded to the maximum stimulus duration of 500 ms.

Traditional Versions of Executive Function Tests

Wisconsin Card Sorting Test (WCST; Heaton et al., 1993)

The WCST is the most well-known test for detecting perseveration and mental rigidity (the ability to form, change, and maintain sets), as well as for examining problem-solving strategies. The test assesses the ability to create and modify categorization principles through the task of classifying a series of cards according to one of three classification criteria (color, shape, and number of elements). Key measures include total errors (perseverative and non-perseverative), perseverative responses, and categories achieved, reflecting success and sorting efficiency. Additional indices include conceptual level responses (correct sequences), attempts to complete the first category, learning to learn (change in error rates), and inability to maintain a set (fewer than nine consecutive correct responses). The Total Achievement score accounts for both the number of categories and the attempts needed to identify them, emphasizing cognitive flexibility and categorization skills (Cianchetti et al., 2005).

Trail Making Test, Form A and B (TMT; Reitan, 1955, 1992; Spreen & Strauss, 1991)

TMT consists of two parts assessing different cognitive functions. Part A evaluates attention, concentration, visual perception, visuospatial processing, and visuomotor skills. It involves connecting numbered circles (1 to 25) in order, without lifting the pen, as quickly as possible. Part B also includes numbered white circles (1 to 13) and additional gray circles (1 to 12). Participants must alternate between connecting the white and gray circles in numerical order, assessing attention, executive functions, and complex conceptual tracking. Both parts measure the time taken to complete the task, with longer times indicating poorer performance.

Verbal Fluency Test (Phonemic and Semantic, see Goodglass & Kaplan, 1983; Lezak, 1995)

The Verbal Fluency Test (Phonemic and Semantic) assesses verbal fluency through 3 phonemic tasks and one semantic task. It measures the number of words produced in a set time. Phonemic fluency involves listing words starting with a specific letter in the Serbian language (e.g., /S/, /K/, /L/) in 60 seconds, excluding repetitions, proper names, and geographical terms. Semantic fluency requires naming as many different animals as possible in one minute, avoiding repetition and irrelevant variations. These tasks also evaluate divergent thinking, as they require the participant to generate multiple solutions rather than one correct answer.

The Advanced Progressive Matrices (APM; Raven et al., 1998)

APM specifically Series II, which includes 36 tasks with a 40-minute time limit, were used to measure general intelligence (g-factor). The Raven Progressive Matrices is a non-verbal, multiple-choice test designed to assess reasoning ability, a key component of the g-factor. Participants identify the missing element that completes a pattern. The total score, derived from the number of correct answers, provides a measure of general intellectual functioning.

Data preparation and analysis

The dataset used for statistical analyses includes three groups of variables: traditional tests (N = 468), Mental Shifting/Flexibility Task (N = 324), and Auditory-Visual Go/No-Go Task (N = 428). All measures (both computerized and traditional tests) were first standardized (z-transformation) to ensure consistent variance across scales and subsequently normalized to approximate a normal data distribution. For normalization, the natural logarithmic function, $\ln(x + 1)$, was used (McElreath, 2020).

Data preparation, calculation of correlations, and estimation of intraclass correlation coefficients (ICC) were conducted using SPSS for Windows v25 (IBM Corp., 2017), while mixed-effects modeling was performed in the R programming environment v4.1.0 (R Core Team, 2020) within the integrated development environment (IDE) RStudio (RStudio, 2023). For ICC estimation, the one-way random effects model was applied following the recommendations of Shrout et al. (1979), with values interpreted based on Cicchetti (1994).

Mixed-effects models were implemented using the “nlme” package in R (Pinheiro et al., 2021) to examine the effect of a single predictor on a single criterion while accounting for the hierarchical structure of the data (i.e., membership within a twin pair). In all analyses, predictors were traditional measures of executive functions, while criterion variables were computer-based tasks. Optimal models for each case were evaluated (based on likelihood ratio tests) by including a random intercept for groups defined by the twin pair identifier, with fixed effects of the predictors compared across groups. Standardized regression coefficients (beta weights) and p-values were calculated for each predictor-criterion pair.

Results

Descriptive statistics for all measures used are presented in Appendix C. These statistics correspond to the previously described transformed scores. Based on the values of skewness and kurtosis, no significant deviations from normal distribution were observed.

Correlations between RTs across different tasks are statistically significant, positive, and of moderate strength. The number of errors across blocks of tasks generally correlates positively, although the relationships are of low intensity. Additionally, correlations between RTs and the number of errors within tasks are predominantly statistically significant, negative, and weak. A summary of these results can be found in Appendices D, E, and F. Reliability ICC values ranged from poor to good (Appendix G).

Mental Shifting/Flexibility Task

The results of the mixed-effects model analyses, where the predictor variables were scores on the traditional tests and the criterion variables were RTs and errors, are presented in Table 1. The most consistent pattern of statistically significant relationships was positive associations with TMT-A, TMT-B, and RTs. Negative relationships were found with the APM and RTs, though these were weaker in magnitude. Most other variable pairs did not exhibit statistically significant relationships. When significant relationships were observed, they were predominantly positive and of low strength.

Auditory-Visual Go/No-Go Task

Similarly, the results of the mixed-effects model analyses for the Auditory-Visual Go/No-Go Task showed positive associations with TMT-A, TMT-B, and RTs, as well as negative and moderate relationships between the APM and RTs, among the most consistent findings (Table 2). The most notable difference between the two tasks was observed in the number of errors for block 3 of the Auditory-Visual Go/No-Go Task. The number of errors from this block exhibited multiple significant (both positive and negative), albeit weak, relationships with other traditional tests. Additionally, statistically significant negative relationships were found only between RTs from the Auditory-Visual Go/No-Go Task and Semantic Fluency.

Table 1

Relationships between traditional tests of executive functions and reaction time and number of errors from the Mental Shifting/Flexibility Task

	B1 RT	B2 RT	B3 RT	B4 RT	B5 RT	B1 Er	B2 Er	B3 Er	B4 Er	B5 Er
Categories completed	-.092	.020	-.088	-.063	-.057	-.047	.062	.002	.039	-.016
Perseverative errors	.067	.042	.107*	.094	.124*	.084	.019	-.029	.011	.073
Perseverative responses	.081	.057	.117*	.103	.133**	.084	.019	-.025	.001	.080
Non-perseverative errors	.026	-.033	.017	.038	.058	.066	.009	-.033	.002	-.004
Failure to maintain set	.105*	-.040	-.092	-.022	-.063	.059	-.058	-.002	.048	.035
Trials to complete the first category	.040	.003	-.042	.015	-.020	-.025	.003	.007	.047	-.011
Total No. of correct answers	.035	-.005	-.048	-.004	-.003	.104*	.010	-.013	.008	.099*
Total No. of errors	.047	.002	.054	.060	.102*	.083	.013	-.041	-.011	.064
Conceptual level responses	-.088	.001	-.091	-.081	-.112	-.054	.034	.023	.004	-.022
Categorizing efficiency	-.084	.002	-.080	-.068	-.072	-.071	.035	.018	.029	-.053
TMT-A reaction time	.163*	.181*	.174*	.181*	.195**	.071	-.056	.013	.057	-.053
TMT-B reaction time	.251**	.261**	.161*	.230**	.168*	.106*	-.010	.024	.083	.023
Phonemic fluency	-.091	-.072	.014	-0.033	-.098	-.015	-.087	.028	-.103*	-.097
Semantic fluency	-.061	-.085	-.051	-.076	-.170**	-.045	-.032	-.039	-.024	-.086
Advanced Progressive Matrices	-.245***	-.199**	-.197**	-.253***	-.181**	-.037	-.012	-.058	.000	.003

Note. B1-B5 = block number; RT = reaction time; Er = number of errors.

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

Table 2

Relationships between traditional tests of executive functions and reaction time and number of errors from the Auditory-Visual Go/No-Go Task

	B1 RT	B2 RT	B3 RT	B1 Er	B2 Er	B3 Er
Categories completed	-.012	.017	-.027	-.020	-.065	-.082
Perseverative errors	.062	.004	.012	.013	.048	.090
Perseverative responses	.075	.006	.010	.023	.047	.096*
Non-perseverative errors	.043	.042	-.036	.058	.046	.110*
Failure to maintain set	-.024	-.071	.010	-.008	-.029	.086
Trials to complete the first category	.014	.007	-.039	-.029	-.045	.043
Total No. of correct answers	-.009	-.028	-.004	-.008	-.034	.065
Total No. of errors	.081	.035	-.015	.037	.047	.093*
Conceptual level responses	-.038	.002	.007	-.005	-.072	-.099*
Categorizing efficiency	-.044	.006	-.011	-.021	-.061	-.094*
TMT-A reaction time	.231**	.129*	.174**	-.044	.038	-.016
TMT-B reaction time	.131*	.134*	.176**	.032	.051	.149*
Phonemic fluency	-.016	-.049	-.058	-.068	-.080	-.026
Semantic fluency	-.107*	-.054	-.128**	-.099*	-.064	-.030
Advanced Progressive Matrices	-.301***	-.148*	-.303***	-.050	-.127*	-.233***

Note. B1-B5 = block number; RT = reaction time; Er = number of errors.

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

Discussion

Traditional paper-and-pencil tests, like the WCST, TMT, and Verbal Fluency tests, have long been used to assess executive functions but are limited in capturing the dynamic and real-time nature of cognitive processes. These tests often suffer from 'task impurity,' where non-EF components, such as visual-spatial processing, also affect performance. Despite these limitations, traditional tests remain valuable due to their well-established psychometric properties, extensive normative data, and ability to assess executive functions in a controlled and reliable manner. They are widely used in clinical settings and have proven effective for diagnosing cognitive impairments. Computer-based tasks have been developed to address the challenges of traditional tests by providing more precise, real-time insights into specific executive functions. However, questions remain about their reliability, validity, and comparison with traditional methods, particularly in clinical and diverse populations.

This study aims to assess the psychometric characteristics of two newly developed computer-based tests for executive functions, specifically focusing on their internal consistency and construct validity. By comparing these tools with traditional methods, we aim to determine their effectiveness in capturing the complexity of EF and their suitability for clinical and research contexts.

In this study, two computer-based tasks were developed to measure two core aspects of EF: inhibition and mental shifting. The Mental Shifting/Flexibility Task showed significant correlations with several traditional executive function measures. Notably, the computer-generated task of mental shifting/flexibility showed a positive correlation with both parts of the TMT, albeit of low intensity. The correlation was somewhat stronger for part B of the TMT, which is expected, as this part assesses more complex conceptual tracking and requires flexibility in shifting mental sets under rapid conceptual transitions (Baron, 2004). Furthermore, the observed correlations between the computer-generated task of mental shifting/flexibility and both TMT parts could be partly explained by the fact that both tasks reflect performance in terms of reaction time, suggesting that general cognitive processing speed might mediate this relationship. However, it is also possible that the connection reflects a more specific capacity for cognitive flexibility and the ability to rapidly switch between tasks or concepts rather than just processing speed.

Additionally, performance on the computer-generated mental shifting task was correlated with measures from the WCST, particularly perseverative responses and errors. Perseverative errors, which occur when a participant continues to use a previously correct sorting strategy despite feedback indicating a change in the rule, are key inverse indicators of cognitive flexibility. These errors typically happen after a rule change, when the participant fails to adapt and reverts to the old category. The correlation between this task and WCST measures suggests that mental shifting ability is linked to the capacity to inhibit previously dominant strategies and adapt to new rules.

The Mental Shifting/Flexibility Task showed correlations with APM, highlighting its connection to general intelligence and problem-solving abilities. A low but significant correlation was found between general cognitive ability and reaction time across all blocks of the task. Such findings suggest that the observed link between these constructs may, at least in part, be attributed to general cognitive processing speed. It could be that individuals with higher intelligence perform faster on shifting and inhibition tasks due to their greater overall cognitive processing speed (Horn & Noll, 1997). Also, the relationship between shifting tasks and general intelligence may reflect the shared demands on cognitive control, such as working memory and attentional control. Both general intelligence and tasks like mental shifting require rapid processing and the ability to reconfigure mental sets and adapt to new rules. Therefore, individuals with higher cognitive ability might be more efficient at deploying these cognitive resources, resulting in faster response times across a variety of tasks (Neubauer et al., 1997; Vernon & Jensen, 1984).

Furthermore, a low but statistically significant correlation was found between the first computer-generated task for mental shifting/flexibility and semantic fluency. This relationship can be explained by the shared cognitive processes between the two tasks. Semantic fluency involves searching semantic memory, switching between concepts, and inhibiting irrelevant responses, all of which require cognitive flexibility and inhibition (Swan & Carmelli, 2002; Schwartz et al., 2003; Troyer et al., 1997). The low correlation suggests that, although both tasks tap into cognitive flexibility, the specific demands—conceptual switching in semantic fluency versus rule-based shifting in the mental shifting task—are not perfectly aligned, which may account for the modest connection between them. In addition to reaction time, two occasional correlations were found with the number of errors in the blocks of the first task and the TMT-B and phonemic fluency tests. These correlations were of very low intensity.

Although the correlations between the first computer-generated task for mental shifting/mental flexibility and traditional executive function measures are statistically significant, they are generally low. These results suggest that the task partially taps into similar cognitive processes such as flexibility, inhibition, and general cognitive ability but does not establish a strong connection with these functions. Furthermore, the observed correlations may be partly explained by general cognitive processing, such as reaction time, rather than by the specific demands of mental flexibility or inhibition.

The second task, the Auditory-Visual Go/No-Go Task, shows a similar pattern of correlations with traditional executive function measures as observed in the previous task. Specifically, reaction time in this task was significantly correlated with both parts of the Trail Making Test, indicating a shared demand for cognitive flexibility, as well as inhibition and processing speed across tasks. A stronger correlation was found for TMT part A, which assesses basic cognitive processing, and part B, which requires more complex task switching and inhibition, reflecting the task's focus on cognitive control. Additionally, performance on the task, taking into account both reaction speed and error rate, was negatively correlated with Raven's Advanced Progressive Matrices scores, suggesting that individuals with higher general cognitive ability tend to perform faster in this task, possibly due to greater processing speed. The negative correlation with semantic fluency further supports the notion that faster reaction times in this inhibition task may reflect greater cognitive efficiency and the ability to switch between tasks or concepts. These findings suggest that inhibition capacity, measured through reaction time, is closely linked to broader cognitive abilities, including processing speed, cognitive flexibility, and general intelligence.

However, no significant correlations were found with more typical inhibition measures from classical tasks, such as those derived from the Wisconsin Card Sorting Test, including Perseverative errors, Perseverative responses, and Failure to maintain a set. This absence of significant associations suggests that while the computer-generated task taps into inhibitory control, it may do so in a slightly different way or engage different cognitive processes compared to traditional inhibition measures.

The lack of significant correlations with measures from classic tasks for assessing executive function inhibition further suggests that these may represent distinct aspects or types of inhibitory function. Research has shown that inhibitory measures derived from different methods (e.g., the Stop-Signal and Stroop tasks) exhibit low intercorrelations - a finding that has been interpreted as evidence for multiple types of inhibitory functions (Khng

& Lee, 2014). Some researchers also argue that while the ability to inhibit a predominant response is central to cognitive control, it remains an open question whether the same neural mechanisms mediate inhibition across various tasks designed to assess it (Wager et al., 2005).

This also applies to other cognitive tasks, as studies have shown that correlations between tasks designed to measure the same executive function are generally low to moderate in intensity (Miyake et al., 2000). One explanation for this is that measures of executive functions obtained from individual tasks are always contaminated by the specific characteristics of the task itself and the stimulus materials used in it.

It is also important to consider that in this task, designed for this study, as well as in other Go/NoGo tasks, there is a certain saturation effect related to working memory and mental shifting. These tasks require the engagement of working memory in order to effectively retrieve and maintain the single rule for when and on which stimuli a response should be inhibited (Luciano et al., 2001).

The correlations between reaction times in different experiments are statistically significant and moderate, while the error rates show low, positive correlations. These findings align with the expected low to moderate phenotypic correlation between mental set shifting and inhibition tasks. Both processes rely on working memory and cognitive flexibility (Diamond, 2013; Friedman & Miyake, 2004), with shifting often requiring the inhibition of previous responses. As such, individuals with better set-shifting abilities are likely to exhibit higher levels of inhibition, as both processes rely on a similar cognitive strategy. The overlap in cognitive resources is supported by the shared involvement of the prefrontal cortex (Wager et al., 2005).

The moderate reaction time correlation suggests that set shifting and inhibition share a cognitive foundation but with subtle differences in cognitive control. Weaker correlations in errors point to individual factors, such as attention or motivation, influencing task performance. Both tasks depend on working memory, highlighting cognitive flexibility as key to managing multiple tasks.

The results of our study showed that the experimental procedures involving reaction time demonstrated moderate to good reliability in most blocks, particularly in the second experiment (Auditory-Visual Go/No-Go

Task). On the other hand, the number of errors in both experiments exhibited poor reliability, suggesting greater variability in participants' responses, which made it more challenging to accurately measure errors in these tasks.

Conclusions and limitations

Validating computer-based assessments is crucial to ensure their reliability and understand their role in evaluating executive functions. The study tested the psychometric properties of these tools, demonstrating that they can complement traditional methods, offering a more comprehensive and flexible approach to understanding cognitive functioning. While the results suggest that the computer-based tasks capture important cognitive abilities such as inhibition and mental shifting, it is clear that further validation is required, particularly in terms of their generalizability across diverse populations and contexts. It will be essential to expand testing to more varied demographic groups, including individuals with neurological disorders, aging populations, and those from different cultural backgrounds, to better understand the broader applicability of these tools.

Moreover, the study highlighted the complexity of linking computer-based assessments with traditional executive function measures. While the low to moderate correlations observed between the computer-based tasks and traditional tests (such as the WCST and TMT) underscore the need for a more nuanced understanding of how different tasks assess executive function, these correlations were expected. This is because computer-based tasks and traditional measures may rely on distinct yet overlapping cognitive processes. Therefore, the predominantly low correlations observed reflect the complex relationship between tasks that share some common cognitive resources but are not fully interchangeable. These correlations are not necessarily a limitation; rather, they emphasize the multidimensional nature of executive function, suggesting that each task captures different facets of cognitive control, which may not always align perfectly.

Additionally, the findings suggest that executive functions are not isolated abilities but interrelated cognitive processes that share common neural resources, such as working memory and cognitive flexibility. This highlights the need for further research to explore the shared and unique contributions of different cognitive processes to performance on various executive function tasks.

Future research should focus on integrating these tools into existing assessment batteries for more accurate insights into cognitive functioning. Optimizing assessments to minimize biases like screen fatigue or technology unfamiliarity is crucial. Additionally, refining task designs to reduce errors from fatigue and working memory saturation is needed. The effectiveness of these tools should be evaluated in diverse populations, including those with neurological disorders, older adults, and people with limited technology experience. Exploring how these tasks interact with other cognitive processes will enhance our understanding of executive functions, while investigating the neurobiological mechanisms could provide deeper insights into both non-clinical and clinical populations.

Given the limitations, the results should be interpreted with caution. The sample, mainly consisting of highly educated individuals, may not represent broader populations, potentially reducing variability in executive function measures. The small sample size and gender imbalance could also limit statistical power. Additionally, consecutively completing the tasks may have led to cognitive fatigue, affecting performance, particularly on tasks requiring sustained attention or inhibitory control. Technology-related biases, such as screen fatigue or unfamiliarity with digital platforms, may have further influenced results.

Another limitation is the lack of established norms for these computer-based tasks, making it difficult to interpret results across populations. Future research should address these gaps, including exploring the applicability of these tools in clinical populations and diverse cultural groups.

Funding

This research was supported by the Science Fund of the Republic of Serbia [#7744418 Genetic and environmental influences on psychological adaptation of children and adults - GENIUS].

Conflict of interest

We have no conflicts of interest to disclose.

Data availability statement

The data supporting the findings of this study will be made available on OSF upon publication, along with instructions for their use.

References

Allport, A., & Wylie, G. (1999). Task-switching: Positive and negative priming of task-set. In G. W. Humphreys, J. Duncan, & A. Treisman (Eds.), *Attention, space, and action: Studies in cognitive neuroscience* (pp. 273–296). Oxford University Press.

Alloway, T. P., & Carpenter, R. K. (2020). The relationship among children's learning disabilities, working memory, and problem behaviours in a classroom setting: Three case studies. *The Educational and Developmental Psychologist,* 37(1), 4–10. https://doi.org/10.1017/edp.2020.1

Baron, I. S. (2004). *Neuropsychological evaluation of the child.* Oxford University Press.

Capovilla, A. G. S., Montiel, J. M., Macedo, E. C., & Charin, S. (2005). *Computerized Stroop Test*. University São Francisco.

Cianchetti, C., Corona, S., Foscoliano, M., Scalas, F., & Sannio–Fancello, G. (2005). Modified Wisconsin Card Sorting Test: Proposal of a supplementary sorting method. *Archives of Clinical Neuropsychology, 20*, 555–558. https://doi.org/10.1016/j.acn.2004.12.002

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Collerton, J., Collerton, D., Yasumichi, A., Barrass, K., Eccles, M., Jagger, C., ... Kirkwood, T. (2007). A comparison of computerized and pencil-and-paper tasks in assessing cognitive function in community-dwelling older people in the Newcastle 85+ study. *Journal of the American Geriatrics Society, 55*, 1630–1635. https://doi.org/10.1111/j.1532-5415.2007.01379.x

Dancy, C. L., & Ritter, F. E. (2017). IGT-Open: An open-source, computerized version of the Iowa Gambling Task. *Behavior Research Methods, 49*(3), 972–978. https://doi.org/10.3758/s13428-016-0759-4

Diamond A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Feenstra, H. E. M., Vermeulen, I. E., Murre, J. M. J., & Schagen, S. B. (2017). Online cognition: Factors facilitating reliable online neuropsychological test results. *Clinical Neuropsychologist, 31*(1), 59–84. https://doi.org/10.1080/13854046.2016.1190405

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, 137(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General, 133*(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

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*, 2208. https://doi.org/10.3389/fpsyg.2018.02208

Garavan, H., Ross, T. J., & Stein, E. A. (1999). Right hemispheric dominance of inhibitory control: an event-related functional MRI study. *Proceedings of the National Academy of Sciences of the United States of America,* 96(14), 8301–8306. https://doi.org/10.1073/pnas.96.14.8301

Golden, C. J. (1978). *Stroop color and word test: a manual for clinical and experimental uses*. Stoelting Company.

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders*. Lea Febiger.

Heaton, R. K. (1981). *Wisconsin card sorting test manual.* Psychological assessment resources.

Heaton, R. K., Chelune, G. J., Tallei, J. L., Key, G. G. & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded*. Psychological Assessment Resources.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2-3), 61-83. https://doi.org/10.1017/S0140525X0999152X

Hinkin, C. H., Castellon, S. A., Hardy, D. J., Granholm, E., & Siegle, G. (1999). Computerized and traditional Stroop Task dysfunction in HIV-1 infection. *Neuropsychology*, 13(2), 306–316. https://doi.org/10.1037//0894-4105.13.2.306

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). The Guilford Press.

IBM Corp. (2017). *IBM SPSS Statistics for Windows, Version 25.0*. IBM Corp.

Jacola, L. M., Willard, V. W., Ashford, J. M., Ogg, R. J., Scoggins, M. A., Jones, M. M., Wu, S., & Conklin, H. M. (2014). Clinical utility of the N-back task in functional neuroimaging studies of working memory. *Journal of Clinical and Experimental Neuropsychology, 36*(8), 875–886. https://doi.org/10.1080/13803395.2014.953039

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychology Review*, 17, 213–233. https://doi.org/10.1007/s11065-007-9040-z

Kane, R. L., & Kay, G. G. (1992). Computerized assessment in neuropsychology: a review of tests and test batteries. *Neuropsychology Review*, *3*(1), 1–117. https://doi.org/10.1007/BF01108787

Kessels, R. P. (2019). Improving precision in neuropsychological assessment: Bridging the gap between classic paper-and-pencil tests and paradigms from cognitive neuroscience. *The Clinical Neuropsychologist*, *33*(2), 357-368. https://doi.org/10.1080/13854046.2018.1518489

Khng, K. H., & Lee, K. (2014). The relationship between Stroop and stop-signal measures of inhibition in adolescents: influences from variations in context and measure estimation. *PloS one*, *9*(7), e101356. https://doi.org/10.1371/journal.pone.0101356

Latendorf, A., Runde, L. M., Salminen, T., & Steinert, A. (2021). Digitization of neuropsychological diagnostics: a pilot study to compare three paper-based and digitized cognitive assessments. *Aging Clinical and Experimental Research*, *33*, 1585-1597. https://doi.org/10.1007/s40520-020-01668-z

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J., & Fischer, J. S. (2004). *Neuropsychological assessment* (4th ed.). Oxford University Press.

Luciano, M., Wright, M., Smith, G. A., Geffen, G. M., Geffen, L. B., & Martin, N. G. (2001). Genetic covariance among measures of information processing speed, working memory, and IQ. *Behavior Genetics*, *31*(6), 581–592. https://doi.org/10.1023/a:1013397428612

McDonald, A. S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments.

Computers & Education, 39, 299–312. https://doi.org/10.1016/S0360-1315(02)00032-5

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman & Hall. https://doi.org/10.1201/9780429029608

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114*(3), 449–458. https://doi.org/10.1037/0033-2909.114.3.449

Miller, J. B., & Barr, W. B. (2017). The technology crisis in neuropsychology. *Archives of Clinical Neuropsychology, 32*(5), 541–554. https://doi.org/10.1093/arclin/acx050

Mitrović, D., Smederevac, S., Delgado-Cruzata, L., Sadiković, S., Pajić, D., Prinz, M., Budimlija, Z., Oljača, M., Kušić-Tišma, J., Vučinić, N., & Milutinović, A. (2024). Personality and COMT gene: Molecular-genetic and epigenetic associations with NEO-PI-R personality domains and facets in monozygotic twins. *Frontiers in Genetics*, 15, Article 1455872. https://doi.org/10.3389/fgene.2024.1455872

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, 49–100. https://doi.org/10.1006/cogp.1999.0734

Morris, R. G., Miotto, E. C., Feigenbaum, J. D., Bullock, P., & Polkey, C. E. (1997). The effect of goal-subgoal conflict on planning ability after frontal- and temporal-lobe lesions in humans. *Neuropsychologia, 35*(8), 1147-1157. https://doi.org/10.1016/S0028-3932(97)00009-2

Neubauer, A. C., Riemann, R., Mayer, R., & Angleitner, A. (1997). Intelligence and reaction times in the Hick, Sternberg and Posner paradigms. *Personality and Individual Differences, 22*(6), 885–894. https://doi.org/10.1016/S0191-8869(97)00003-2

Park, S. Y., & Schott, N. (2022). The trail-making-test: Comparison between paper-and-pencil and computerized versions in young and healthy older adults. *Applied Neuropsychology. Adult*, *29*(5), 1208–1220. https://doi.org/10.1080/23279095.2020.1864374

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2021). *Nlme: Linear and Nonlinear Mixed Effects Models* (Version 3.1-153). R package. https://CRAN.R-project.org/package=nlme

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual*: Section 1, general overview, 1998 edition. Oxford Psychologists Press Ltd.

Reitan, R. M. (1955). The relation of the trail making test to organic brain damage. *Journal of Consulting Psycholog*y, 19, 393–394. https://doi.org/10.1037/h0044509

Riordan, P., Lombardo, T., & Schulenberg, S. E. (2013). Evaluation of a computer-based administration of the Rey complex figure test. *Applied Neuropsychology, 20*(3), 169–178. https://doi.org/10.1080/09084282.2012.670171

RStudio Team. (2023). *RStudio: Integrated Development Environment for R* (Version 2023.12.0 Build 369). RStudio, PBC. https://posit.co/

Sanders, L. M. J., Hortobágyi, T., Balasingham, M., Van der Zee, E. A., & van Heuvelen, M. J. G. (2018). Psychometric properties of a Flanker Task in a sample of patients with dementia: A pilot study. *Dementia and Geriatric Cognitive Disorders Extra, 8*(3), 382–392. https://doi.org/10.1159/000493750

Schatz, P., & Browndyke, J. (2002). Applications of computer-based neuropsychological assessment. *The Journal of Head Trauma Rehabilitation*, *17*(5), 395–410. https://doi.org/10.1097/00001199-200210000-00003

Schwartz, S., Baldo, J., Graves, R. E., & Brugger, P. (2003). Pervasive influence of semantics in letter and category fluency: A multidimensional approach. *Brain and Language, 87*(3), 400–411. https://doi.org/10.1016/S0093-934X(03)00141-X

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Smederevac, S., Mitrović, D., Sadiković, S., Milovanović, I., Branovački, B., Dinić, B. M., & others. (2019). Serbian twin registry. *Twin Research and Human* Genetics, 22(6), 660–666. https://doi.org/10.1017/thg.2019.114

Spreen, O., & Strauss, E. (1991). *A compendium of neuropsychological tests*. Administration, norms and commentary. Oxford University Press.

Steinmetz, J. P., Brunner, M., Loarer, E., & Houssemand, C. (2010). Incomplete psychometric equivalence of scores obtained on the manual and the computer version of the Wisconsin Card Sorting test? *Psychological*

Assessment, 22(1), 199–202. https://psycnet.apa.org/doi/10.1037/a0017661

Swan, G. E., & Carmelli, D. (2002). Evidence for genetic mediation of executive control: a study of aging male twins. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences,* 57, 133–143. https://doi.org/10.1093/geronb/57.2.P133

Tien, A. Y., Spevack, T. V., Jones, D. W., Pearlson, G. D., Schlaepfer, T. E., & Strauss, M. E. (1996). Computerized Wisconsin Card Sorting Test: Comparison with manual administration. *Kaohsiung Journal of Medical Sciences,* 12, 479–485.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology, 11*(1), 138–146. https://doi.org/10.1037/0894-4105.11.1.138

Tyburski, E., Kerestey, M., Kerestey, P., Radoń, S., & Mueller, S. T. (2021). Assessment of motor planning and inhibition performance in non-clinical sample—reliability and factor structure of the Tower of London and Go/No Go computerized tasks. *Brain Sciences, 11*(11), 1420. https://doi.org/10.3390/brainsci11111420

Vernon, P. A., & Jensen, A. R. (1984). Individual and group differences in intelligence and speed of information processing. *Personality and Individual Differences, 5*(4), 411–423. https://doi.org/10.1016/0191-8869(84)90006-0

Welsh, M. C., Cicerello, A., Cuneo, K., & Brennan, M. (1995). Error and temporal patterns in Tower of Hanoi performance: Cognitive mechanisms and individual differences. *Journal of General Psychology, 122*(1), 69-81. https://doi.org/10.1080/00221309.1995.9914919

Vermeent, S., Dotsch, R., Schmand, B., Klaming, L., Miller, J. B., & van Elswijk, G. (2020). Evidence of validity for a newly developed digital cognitive test battery. *Frontiers in Psychology, 11*, 770. https://doi.org/10.3389/fpsyg.2020.00770

Wager, T. D., Sylvester, C. Y., Lacey, S. C., Nee, D. E., Franklin, M., & Jonides, J. (2005). Common and unique components of response inhibition revealed by fMRI. *NeuroImage*, *27*(2), 323–340. https://doi.org/10.1016/j.neuroimage.2005.01.054

Wagner, G. P., & Trentini, C. M. (2009). Assessing executive functions in older adults: a comparison between the manual and the computer-based versions of the Wisconsin Card Sorting Test. *Psychology & Neuroscience*, *2*, 195-198. https://doi.org/10.3922/j.psns.2009.2.011

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology, 7*(2), 131-149. https://doi.org/10.1080/87565649109540577

Appendix

Appendix A

Table A1

Sociodemographic characteristics of study subjects

Age (years; *M*, *SD*)	Range (18-58) 24.06 ± 7.02
Gender (*N*; %)	
male	134 (28.63%)
female	334 (71.37%)
Education level (N; %)	
Primary education (8 years in total)	2 (0.87%)
Secondary (11-12 years in total)	59 (25.76%)
Higher School and University (16-17 years in total)	50 (21.83%)
Student	113 (49.34)%
Other	5 (2.18%)

Note. *N* - number of participants; Min - minimum age; Max - maximum age; *M* - mean age; *SD* - standard deviation of age.

Appendix B

Table B1

Twin sample description

	N	%	Min	Max	*M*	*SD*
MZ pairs	153	65.38	18	58	24.62	7.38
DZ pairs	81	34.62	18	48	23.02	6.24
MZ male pairs	37	24.18	18	58	23.69	7.41
MZ female pairs	116	75.82	18	47	24.91	7.37
SS DZ pairs	45	55.56	18	48	23.27	6.48
DS DZ pairs	36	44.44	18	41	22.72	6.00
Male participants	138	29.49	17	58	22.98	6.38
Female participants	330	70.51	16	48	24.52	7.24

Note. *N* - number of pairs/participants; *Min* - minimum age; *Max* - maximum age; *M* - mean age; *SD* - standard deviation of age; All descriptive parameters correspond to the pairs or participants specified in the first column of the table; MZ – monozygotic; DZ – dizygotic; SS – same sex; DS – different sex.

Appendix C

Table C1

Descriptive statistical parameters for all used measures

Variable	Min	Max	*M*	*SD*	Skewness	Kurtosis
Categories completed	1.50	238.00	119.36	46.29	-1.66	1.06
Perseverative errors	1.00	238.00	118.91	68.81	0.03	-1.21
Perseverative responses	1.00	238.00	118.91	68.83	0.03	-1.22
Non-perseverative errors	1.00	238.00	118.83	68.82	0.03	-1.21
Failure to maintain set	76.00	238.00	119.22	56.78	0.86	-1.05
Trials to complete the first category	1.50	238.00	119.88	65.96	0.14	-1.09
Total No. of correct answers	1.00	238.00	118.50	68.41	0.02	-1.20
Total No. of errors	1.00	238.00	118.83	68.98	0.02	-1.21
Conceptual level responses	7.81	96.00	71.95	17.13	-1.20	0.98
Categorizing efficiency	0.00	99.00	68.39	26.96	-0.99	-0.33
TMT-A reaction time	1.00	238.00	119.00	68.77	0.02	-1.20
TMT-B reaction time	1.00	238.00	119.81	68.76	-0.01	-1.20
Phonemic fluency	3.00	22.33	11.43	3.16	0.21	-0.08
Semantic fluency	10.00	44.00	24.04	5.53	0.49	0.51
Advanced Progressive Matrices	2.00	35.00	20.72	6.17	-0.44	0.05
T1 B1 RT	-2.74	2.74	0.00	1.00	0.00	-0.10
T1 B2 RT	-2.74	2.74	0.00	1.00	0.00	-0.10
T1 B3 RT	-2.74	2.74	0.00	1.00	0.00	-0.10
T1 B4 RT	-2.74	2.74	0.00	1.00	0.00	-0.10
T1 B5 RT	-2.74	2.74	0.00	1.00	0.00	-0.10

T1 B1 Er	-2.50	2.74	0.05	0.87	0.50	0.37
T1 B2 Er	-1.20	2.74	0.03	0.91	0.39	-0.41
T1 B3 Er	-0.63	2.74	0.06	0.81	0.94	-0.07
T1 B4 Er	-0.81	2.74	0.05	0.86	0.68	-0.30
T1 B5 Er	-2.25	2.74	0.04	0.87	0.50	-0.05
T2 B1 RT	-2.83	2.83	0.00	1.00	0.00	-0.08
T2 B2 RT	-2.83	2.83	0.00	1.00	0.00	-0.08
T2 B3 RT	-2.83	2.83	0.00	1.00	0.00	-0.08
T2 B1 Er	-0.75	2.83	0.05	0.85	0.76	-0.24
T2 B2 Er	-0.98	2.83	0.04	0.89	0.54	-0.41
T2 B3 Er	-1.81	2.83	0.01	0.97	0.12	-0.34

Note. The number of participants for the cognitive measures was $N = 468$; T1 = Mental Shifting/Flexibility Task ($N = 324$); T2 = Auditory-Visual Go/No-Go Task ($N = 428$); B1-B5 = block number; RT = reaction time; Er = number of errors.

Appendix D

Table D1

Correlations between RTs and the number of errors of computer-based tasks

	T2 B1 RT	T2 B2 RT	T2 B3 RT	T2 B1 Er	T2 B2 Er	T2 B3 Er
T1 B1 RT	.306**	.288**	.312**	-.072	-.075	-.041
T1 B2 RT	.414**	.324**	.311**	-.124*	-.115*	.118*
T1 B3 RT	.454**	.445**	.359**	-.122*	-.171**	.017
T1 B4 RT	.473**	.412**	.389**	-.188**	-.091	-.026
T1 B5 RT	.450**	.383**	.391**	-.158**	-.070	-.002
T1 B6 RT	.460**	.397**	.318**	-.187**	-.118*	.024
T1 B1 Er	.126*	.035	.111*	.015	.114*	.109
T1 B2 Er	-.096	-.064	-.075	.139*	.028	.088
T1 B3 Er	-.173**	-.104	-.100	.152**	.159**	.042
T1 B4 Er	-.135*	-.127*	.083	.134*	.169**	.142*
T1 B5 Er	-.175**	-.184**	-.107	.151**	.175**	.029
T1 B6 Er	-.066	-.071	.033	.032	.104	.182**

Note. T1 = Mental Shifting/Flexibility Task (*N* = 324); T2 = Auditory-Visual Go/No-Go Task (*N* = 428); B1-B5 = block number; RT = reaction time; Er = number of errors.

Appendix E

Table E1

Correlations between reaction times (RTs) and the number of errors in the same computer-based task (Mental Shifting/Flexibility Task)

	E1 B1 Er	E1 B2 Er	E1 B3 Er	E1 B4 Er	E1 B5 Er
E1 B1 RT	.215**	-.244**	.004	-.210**	-.094
E1 B2 RT	.085	-.089	-.020	-.178**	-.081
E1 B3 RT	-.087	-.231**	-.124*	-.293**	-.178**
E1 B4 RT	-.022	-.149**	-.013	-.208**	-.169**
E1 B5 RT	-.042	-.143**	-.035	-.250**	-.144**

Note. B1-B5 = block number; RT = reaction time; Er = number of errors.

Appendix F

Table F1

Correlations between reaction times (RTs) and the number of errors in the same computer-based task (Auditory-Visual Go/No-Go Task)

	E2 B1 Er	E2 B2 Er	E2 B3 Er
E2 B1 RT	-.249**	-.122*	.003
E2 B2 RT	-.232**	-.301**	-.070
E2 B3 RT	-.149**	.050	.003

Note. B1-B3 = block number; RT = reaction time; Er = number of errors.

Appendix G

Table G1

Intraclass correlation coefficients (ICC) for used measures

Variable	ICC	*p* value	Evaluation of the reliability
Categories completed	0.330	0.001	Poor
Perseverative errors	0.276	0.007	Poor
Perseverative responses	0.278	0.007	Poor
Non-perseverative errors	0.254	0.013	Poor
Failure to maintain set	0.099	0.214	Poor
Trials to complete the first category	0.202	0.042	Poor
Total No. of correct answers	0.084	0.251	Poor
Total No. of errors	0.256	0.012	Poor
Conceptual level responses	0.247	0.016	Poor
Categorizing efficiency	0.289	0.005	Poor
TMT-A reaction time	0.533	0.000	Moderate
TMT-B reaction time	0.483	0.000	Moderate
Phonemic fluency	0.596	0.000	Good
Semantic fluency	0.543	0.000	Moderate
Advanced Progressive Matrices	0.772	0.000	Excellent
T1 B1 RT	0.584	0.000	Moderate
T1 B2 RT	0.603	0.000	Good
T1 B3 RT	0.574	0.000	Moderate
T1 B4 RT	0.652	0.000	Good
T1 B5 RT	0.682	0.000	Good
T1 B1 Er	0.380	0.001	Poor
T1 B2 Er	-0.005	0.513	Poor

T1 B3 Er	0.160	0.134	Poor
T1 B4 Er	0.525	0.000	Moderate
T1 B5 Er	0.080	0.297	Poor
T2 B1 RT	0.639	0.000	Good
T2 B2 RT	0.511	0.000	Moderate
T2 B3 RT	0.581	0.000	Moderate
T2 B1 Er	0.249	0.019	Poor
T2 B2 Er	0.359	0.001	Poor
T2 B3 Er	0.314	0.003	Poor

Note. T1 = Mental Shifting/Flexibility Task (*N* = 324); T2 = Auditory-Visual Go/No-Go Task (*N* = 428); B1-B5 = block number; RT = reaction time; Er = number of errors; ICC values were interpreted as follows (Cicchetti, 1994): values below 0.40 indicate poor reliability, values between 0.40 and 0.59 indicate moderate reliability, values between 0.60 and 0.74 indicate good reliability, and values of 0.75 or higher indicate excellent reliability.