

# UC Davis

## UC Davis Previously Published Works

### Title

NIH Toolbox Cognition Battery (NIHTB-CB): List Sorting Test to Measure Working Memory

### Permalink

<https://escholarship.org/uc/item/0kj27902>

### Journal

Journal of the International Neuropsychological Society, 20(6)

### ISSN

1355-6177

### Authors

Tulsky, David S  
Carlozzi, Noelle  
Chiaravalloti, Nancy D  
[et al.](#)

### Publication Date

2014-07-01

### DOI

10.1017/s135561771400040x

Peer reviewed



# HHS Public Access

Author manuscript

*J Int Neuropsychol Soc.* Author manuscript; available in PMC 2015 July 01.

Published in final edited form as:

*J Int Neuropsychol Soc.* 2014 July ; 20(6): 599–610. doi:10.1017/S135561771400040X.

## NIH Toolbox Cognition Battery (NIHTB-CB): The List Sorting Test to Measure Working Memory

David S. Tulsky<sup>1,2</sup>, Noelle Carlozzi<sup>3</sup>, Nancy D. Chiaravalloti<sup>2</sup>, Jennifer L. Beaumont<sup>4</sup>, Pamela A. Kisala<sup>5</sup>, Dan Mungas<sup>6</sup>, Kevin Conway<sup>7</sup>, and Richard Gershon<sup>4</sup>

<sup>1</sup>Departments of Rehabilitation Medicine, Orthopedic Surgery, and General Medicine, New York University Langone Medical Center, New York, New York <sup>2</sup>Kessler Foundation Research Center, West Orange, New Jersey <sup>3</sup>Department of Physical Medicine and Rehabilitation, University of Michigan, Ann Arbor, Michigan <sup>4</sup>Department of Medical Social Sciences, Northwestern University, Chicago, Illinois <sup>5</sup>Department of Rehabilitation Medicine, New York University Langone Medical Center, New York, New York <sup>6</sup>Department of Neurology, University of California, Davis, California <sup>7</sup>National Institute on Drug Abuse, Rockville, Maryland

### Abstract

The List Sorting Working Memory Test was designed to assess working memory (WM) as part of the NIH Toolbox Cognition Battery. List Sorting is a sequencing task requiring children and adults to sort and sequence stimuli that are presented visually and auditorily. Validation data are presented for 268 participants ages 20 to 85 years. A subset of participants (N=89) was retested 7 to 21 days later. As expected, the List Sorting Test had moderately high correlations with other measures of working memory and executive functioning (convergent validity) but a low correlation with a test of receptive vocabulary (discriminant validity). Furthermore, List Sorting demonstrates expected changes over the age span and has excellent test-retest reliability. Collectively, these results provide initial support the construct validity of the List Sorting Working Memory Measure as a measure of working memory. However, the relation between the List Sorting Test and general executive function has yet to be determined.

### Keywords

Cognition; Working Memory; Memory; Short-Term; Neuropsychological Tests; Executive Function; Adult

## INTRODUCTION

The National Institutes of Health (NIH) has recognized that the lack of continuity of outcomes measures across the research portfolio hinders the ability to aggregate findings

Correspondence may be addressed to: David S. Tulsky, Assessment Research and Translation, Rusk Institute/Department of Rehabilitation Medicine, New York University Langone Medical Center, Ambulatory Care Center 240 E. 38<sup>th</sup> St., 17<sup>th</sup> Floor, New York, NY 10016. david.tulsky@nyumc.org; Phone: 212-263-6164; fax: 212.263.2683.

The authors have nothing to disclose.

across studies and populations. NIH sought to improve the “economies of scale in the research enterprise” by having common variables that could be analyzed across studies (Hodes, Insel, & Landis, 2013). Therefore, the NIH Toolbox project was initiated to develop a set of state-of-the-art outcomes measures to serve as common data elements for the evaluation of cognitive, social, emotional and motor functioning across the lifespan (Gershon et al., 2010; Gershon et al., 2013). The Cognitive Health Domain consists of discrete subtests that have been designed to evaluate Working Memory, Executive Function, Processing Speed, Episodic Memory, and Language, respectively, in a total of less than 30 minutes (Weintraub et al., 2013). This manuscript describes the validation of the new measure of working memory, the List Sorting Test, in adults without known neurological deficits. The development and validation of the List Sorting Test for use in children has been previously described by Tulsky and colleagues (Tulsky et al., 2013).

The term *working memory* (WM) was first introduced by Newell and Simon (Newell, 1973; Newell & Simon, 1972) but is now more closely associated with Alan Baddeley and his colleagues (Baddeley, 1987, 1992, 2002, 2010; Baddeley & Hitch, 1974). WM commonly refers to the cognitive process that stores and manipulates a limited amount of information over a brief period of time (Baddeley, 1987, 1992). It is essentially an updated conceptualization of “short-term memory,” a passive storage buffer that became overloaded when the amount of information exceeded its capacity (Miller, 1956). WM refers to the ability to 1) process information across tasks and modalities, 2) hold information in a short-term store, 3) manipulate information, and 4) hold the products of that manipulation in the same short-term store. The working memory conceptual workspace is an active system that forms a strong basis for information processing overall (Baddeley, 2002).

In the original conceptualization, WM was comprised of a central executive system (CES) and two “slave” systems that contain the phonological loop (PL) and the visuospatial sketchpad (VS) (Baddeley, 1986; Baddeley & Hitch, 1974; Logie, 1996). The CES is the most important but least understood component of the working memory model (Baddeley, 1996), theorized to have no retention capacity but providing a set of mechanisms to control all the WM processes through its attentional capacity. The CES is thus thought to be responsible for the selection, initiation, and termination of processing routines within working memory. Baddeley (1996) described the CES as a “homunculus” because it is such a powerful system that is closely related to executive functioning. According to the classic model postulated by Baddeley, the CES functions include the coordination of the two slave systems and the manipulation of their contents. The CES is also responsible for the control of encoding and retrieval strategies, as well as the focus, dividing and switching of attention. In contrast, the slave systems are conceptualized to be responsible for the temporary maintenance of information; the phonological loop is where verbal material is stored and processed while the visuospatial sketch pad is where spatial material is stored and processed. In the face of empirical findings that did not support a 3 component system of working memory, Baddeley (2000) updated the model, introducing a fourth component: the episodic buffer, which is a limited capacity system that provides passive and temporary storage of information. The episodic buffer provides an interface between the above subsystems and information from perception and long term memory (Baddeley, 2000, 2001, 2010). The

Baddeley model has been widely and successfully applied in basic and applied psychology, as well as in neuroscience (Miyake & Shah, 1999; Repovs & Bresjanak, 2006).

WM is important for a wide variety of cognitive skills, such as long-term memory (Jonides, 1995), learning (Kirasic, Allen, Dobson, & Binder, 1996), reasoning (Salthouse, Mitchell, Skovronek, & Babcock, 1989), problem solving, planning, and active listening (Jonides, 1995). Miyake and colleagues (2000) have identified WM as an “updating and monitoring” component of executive functions that play a central role in WM performance, since it manipulates WM content in order to select, from time to time, the goal relevant ones. McCabe and colleagues (2010) specifically examined the relationship between WM and executive functioning, concluding the 2 constructs to share a factor of “executive attention”, which predicts higher level cognitive processes such as episodic memory. Indeed, WM is necessary for common yet complex everyday activities which require multitasking. For instance, people will often take notes while talking on the telephone at work, perform mental arithmetic at school or compare and calculate prices at a store. Individuals with significant WM deficits have great difficulty processing and learning information (Smith, Jonides, & Koeppe, 1996) and basic everyday tasks like writing and reading comprehension may be affected by impairments in WM (Engle, 1996). As such, WM is a critical and primary component of cognition, and WM disruption can be particularly problematic to everyday functioning.

The importance of WM in cognition throughout the lifespan is evident when one considers that published studies examine WM from infancy through older adulthood (Pelphrey & Reznick, 2003). Structural components of WM are thought to be in place by age 4 (Alloway, Gathercole, & Pickering, 2006) and much work has been done with pediatric populations (Conlin, Gathercole, & Adams, 2005). However, the vast majority of studies have involved school-age children and beyond (e.g., 6 years), likely due to the ease of communication and ability to follow instructions in children ages 6 and over. Studies have demonstrated that WM improves significantly in conjunction with overall cognitive development, with WM span doubling in capacity between the ages of 5 and 10 (Riggs, McTaggart, & Simpson, 2006) and tripling by the time the child reaches young adulthood (Dempster, 1981). WM integrity has been linked to scholastic development (Hitch, Towse, & Hutton, 2001) and letter knowledge (de Jong & Olson, 2004), and its impairment has been linked to reading disabilities (de Jong, 1998). WM remains relatively stable throughout adulthood and reduced WM capacity is evident in older adults. Many have posited that declines in cognitive functions will often result in poor WM performance. For instance, Tim Salthouse (1994) attributes WM decline in older adulthood to decreased processing speed, rather than to changes in WM itself. Other explanations of age-related WM decline have been explained by a heightened distractibility and deficits in inhibiting irrelevant information in older adults rather than direct deficits in WM (Hasher, Stolfus, Zacks, & Rypma, 1991).

Impaired WM is common in medical disorders with neurological involvement such as Multiple Sclerosis (Grigsby, Ayarbe, Kravcisin, & Busenbark, 1994; Grigsby, Busenbark, Kravcisin, Kennedy, & Taylor, 1994), Traumatic Brain Injury (McAllister, Flashman, Sparling, & Saykin, 2004; Willmott, Ponsford, Hocking, & Schönberger, 2009), HIV (Woods, Moore, Weber, & Grant, 2009) and Schizophrenia (Van Snellenberg, 2009). Such

WM deficits have been documented both behaviorally (Demaree, DeLuca, Gaudino, & Diamond, 1999) and in neuroimaging studies (Chiaravalloti et al., 2005; McAllister et al., 2004). Additionally, such deficits have been shown to impact higher order cognitive processes, such as long term memory (Litvan, Grafman, Vendrell, & Martinez, 1988; Rao et al., 1993) and executive control (Orellana & Slachevsky, 2013). However, debate remains as to whether these difficulties arise from impairment in the CES (D'Esposito et al., 1996) or from within the PL (Litvan et al., 1988; Rao et al., 1993).

Given the prevalence of WM dysfunction in neurological populations, it is not surprising that various neuroimaging techniques have been applied to examine the neurofunctional basis of these deficits. Since the inception of neuroimaging, WM has received considerable attention in healthy individuals in an effort to understand its anatomical and functional architecture (Braver et al., 1997; Fuster, 1989; Goldman-Rakic, 1987; Jacobsen, 1935; Manoach et al., 1997; Milner, 1964). Such studies have revealed a distributed network of cerebral involvement during WM functioning, including activation of the prefrontal and premotor regions of the frontal lobes (e.g., the middle and inferior frontal gyri) (Belger et al., 1998; Braver et al., 1997; Courtney, Ungerleider, Keil, & Haxby, 1997; Crosson et al., 1999; Curtis & D'Esposito, 2003; Grossman et al., 1994; Linden, 2007) as well as parietal (Braver et al., 1997; McAllister et al., 1999; Paulesu, Frith, & Frackowiak, 1993) and temporal activation (Paulesu et al., 1993; Salmon et al., 1996; Seidman et al., 1998). Some regions have been shown to be modality-specific, responsible for verbal working memory in particular (Curtis & D'Esposito, 2003; Linden, 2007). Other regions appear to be multimodal in nature (Curtis & D'Esposito, 2003; Klingberg, Kawashima, & Roland, 1996; Linden, 2007). Most research in healthy individuals has supported the Baddeley WM model through anatomic differentiation between the cerebral substrates of maintenance (PL and VS) and manipulation (CES). Findings indicate that simple maintenance of information is associated with ventral prefrontal regions (Awh et al., 1996). The dorsal prefrontal cortex becomes involved when information needs to be manipulated (Beatty et al., 1996; D'Esposito et al., 1998; Owen, Evans, & Petrides, 1996), which is thought to reflect central executive system (CES) recruitment.

WM capacity has also been linked to differential patterns of cerebral activation during WM task performance across the lifespan. Results indicate that the development of adult cognition is reliant upon increased cerebral activation in localized regions necessary for effective WM performance (Scherf, Sweeney, & Luna, 2006). Specifically, young children have shown limited activation of core WM regions (e.g. dorsolateral prefrontal cortex and parietal regions), with greater activation documented in these regions in adolescents (Klingberg, Forssberg, & Westerberg, 2002). Adults demonstrate the most specialized network, including left dorsolateral prefrontal cortex, ventrolateral prefrontal cortex and the supramarginal gyrus (Scherf, Sweeney, & Luna, 2006). These changes in cerebral activation patterns have shown functional relevance, with WM capacity being positively correlated with such activation (Klingberg, Forssberg, & Westerberg, 2002; Crone, Wendelken, Donohue, van Leijenhorst, & Bunge, 2006; Kwon, Reiss, & Menon, 2002). In fact, neural network models have indicated that the strength of frontoparietal connectivity underlies WM capacity (Edin et al., 2009; Edin, Macoveanu, Olesen, Tegner, & Klingberg, 2007).

WM is frequently assessed with tasks requiring simultaneous storage and processing, such as complex span tasks (e.g., backward digit span, letter-number sequencing, etc.) and various updating tasks (e.g., n-back, keeping track, etc.) (Conway et al., 2005; Schmiedek, Hildebrandt, Lovden, Lindenberger, & Wilhelm, 2009). WM is significantly related to other tasks of sustained attention, executive functioning, and processing speed. Kyllonen & Christal (1990) demonstrated that WM was related to reasoning tests and discussed that WM was highly related to general intelligence. Miyake et al (2000) showed that working memory, while separate from other forms of executive function, is related to set shifting and inhibition and shares an underlying commonality. Similarly, Lehto (1996) demonstrated that WM tasks are correlated with the Wisconsin Card Sorting Tasks or set shifting tasks. WM abilities and executive functioning have been noted to be similarly impacted in populations such as adults with Huntington's Disease (You et al., 2013) and children with specific language impairment (Vugs, Hendriks, Cuperus, & Verhoeven, 2014), with some authors discussing WM as a facet of executive functioning (Hofmann, Schmeichel, & Baddeley, 2012). Given the importance of WM to overall cognitive development and the execution of higher order cognitive tasks, in the NIH Toolbox, emphasis and concern was placed upon the selection of a valid and reliable task that could be utilized across the lifespan and that could be administered quickly as a core component of a 30 minute battery.

As reported by Tulsky et al. (2013), the Toolbox List Sorting Working Memory Test is a sequencing task requiring children and adults to sort and sequence information. Participants are presented with a series of stimuli (i.e. illustrated pictures of an animal or a piece of food), each of which is both visually and auditorily presented by computer (see Figure 1). A picture of each stimulus is displayed on the computer monitor for 2 seconds while the name of the stimulus is simultaneously being read via a computerized voice; stimuli are presented one after another, in a seamless fashion. The examinee is required to remember each stimulus in a series, mentally reorder them from smallest to largest, and recite the names of the stimuli in this order. The final version of this task included both a 1-list and 2-list component. In the "1-list" section, only one type of item stimulus is presented (e.g., "animals" or "food"). Participants begin by sequencing a two-item string and with each correct response, the string is increased by a single item (up to a maximum of a seven-item string). If the participant is unable to sequence the string correctly, they are provided with a second trial of the same number of items in the string; the task is discontinued when the participant provides incorrect responses on two trials with the same number of items in the string or when the participant correctly sequences all seven items. Participants begin with the 1-list version of the task (which requires sequencing among a single category, either "animals" or "food"), and then, upon discontinuing the 1-list section, move on to the 2-list section (which requires sequencing of both "food" and "animals" within a string). In the 2-list section, the examinee is required to sort the stimuli by category prior to sequencing the stimuli in size order, requiring the examinee to both sort and dual sequence the information. Participants begin the 2-list section with a two-item string, which is increased by a single item with each correct response (also up to a maximum of a seven-item string). They are provided a second trial of the same number of items in the string following an incorrect response and the task is discontinued when the participant provides incorrect responses on two trials with the same number of items or when the participant correctly sequences all

seven items. List Sorting scores are based upon a sum of the total correct across both lists which comprise the List Sorting “Total Score.” The raw sum score is then transformed to a standardized t-metric (mean=50, SD=10).

The NIH Toolbox Cognition team chose a sequencing task because such a task had proven successful in measuring WM in the Wechsler Adult Scales of Intelligence, Third Edition (Tulsky, Saklofske, & Zhu, 2003; Wechsler, 1997), in individuals with schizophrenia (Gold, Carpenter, Randolph, Goldberg, & Weinberger, 1997), and in older adults (Mungas, Reed, Marshall, & Gonzalez, 2000). The team developed a task that would be valid for the youngest children that was modeled from the sequence test in the Spanish and English Neuropsychological Assessment Scales (Mungas, Reed, Marshall, & Gonzalez, 2000), in which animals and/or fruit were verbally presented to examinees who were instructed to re-order the stimuli in size order.

Extensive pilot development was conducted to help mold the test into a task that could be completed by very young children (Tulsky et al., 2013). However, a small percentage of children 6 did not complete the first items correctly on either the 1-list or 2-list tasks, prompting the team to remove it as part of the early childhood battery. The test is recommended as part of the core Toolbox battery with children 7. The administration instructions, starting points, and discontinue rules are identical for children and adults on both the 1-list item and 2-list items. Since information is presented both visually and auditorily, scores do not simply reflect one subcomponent (e.g., PL or VS) of WM. Instead, the List Sorting Test is designed to globally assess the WM system.

## Method

### Participants

A sample of 268 participants ages 18 to 85 years (n=159 ages 20–60 and n=109 ages 65 to 85) was recruited from four testing sites across the United States. Participants ages 20–60 were recruited with flyers advertising a research study for healthy volunteers. For the older adults (age group 65–85), 62 of the participants were recruited from a pool of known cognitively healthy individuals participating in other NIH studies while the remaining 47 participants were recruited through the use of flyers. No formal screening of health or cognitive functioning was performed, but self-report questionnaires provided information about health status. Study coordinators established and monitored a stratification matrix so that key demographic variables (i.e., age, education level, gender, and ethnicity) could be distributed. A subset of 89 participants (approximately 33%) completed a retest 7 to 21 days later to assess test-retest reliability. The participants in the validation phase are described in detail in Weintraub et al (2013). The institutional review board at each collaborating site reviewed and approved this study, and all study procedures were carried out in accordance with the principles outlined by the Helsinki Declaration.

### Measures

Participants were tested with the Toolbox List Sorting Working Memory Test and criterion measures including the Wechsler Adult Intelligence Scale, 4<sup>th</sup> Edition Letter-Number



Sequencing subtest and the Paced Auditory Serial Addition Test (Gronwall, 1977) (that would provide indices of concurrent validity) and the Peabody Picture Vocabulary Test, 4<sup>th</sup> Edition (that would provide an index of discriminant validity). Other criterion measures of executive functioning (e.g., Delis Kaplan Executive Function System Color-Word Test) and processing speed (e.g., the Wechsler Adult Intelligence Scale, Fourth Edition Processing Speed Composite) were expected to be related to WM measures. Finally, other NIH Toolbox measures were administered (in particular, the Dimensional Change Card Sorting and the Flanker tests) which are described in other manuscripts in this special issue (Zelazo et al., 2013).

**The Toolbox List Sorting Working Memory Test**—As described previously, the List Sorting Test involves size order sequencing of familiar stimuli. In the validation version, the task was discontinued when two trials of the same length were failed. This task takes approximately 10 minutes to administer. Test scores consisted of combined total items correct on the one- and two-list versions of the task (maximum 28).

### Validation Measures

**WAIS-IV Letter-Number Sequencing (Wechsler, 2008):** In this test, participants are presented with a mixed list of numbers and letters and asked to repeat the list by saying the numbers first in ascending order and then the letters in alphabetical order. This subtest has a strong WM component (Crowe, 2000; Gold, Carpenter, Randolph, Goldberg, & Weinberger, 1997; Haut, Kuwabara, Leach, & Arias, 2000) and has demonstrated discrimination between individuals with TBI and matched controls (Donders, Tulsky, & Zhu, 2001). Scores reflect the number of correct responses (maximum 30).

**WAIS-IV Coding (Wechsler, 2008):** This test requires the participant to associate numbers and symbols using a key and has demonstrated sensitivity to motor coordination, short-term memory deficits, and visuo-perceptual problems (Tulsky et al., 2003). Scores reflect number correct in 120 seconds (maximum 135).

**WAIS-IV Symbol Search (Wechsler, 2008):** This subtest is designed to measure speed at processing new information. Participants are presented with a series of paired groups; each pair consists of a target group (two symbols) and a search group (five symbols). The participant must identify whether either target symbol is part of the search group. Scores reflect number correct minus number incorrect in 120 seconds (maximum 60).

**WAIS-IV Processing Speed Index (PSI) (Wechsler, 2008):** PSI is a composite score that combines performance on WAIS-IV coding and WAIS-IV Symbol Search. PSI scores are significantly lower in individuals with a number of neuropsychological and psychoeducational problems (Hawkins, 1998; Martin, Donders, & Thompson, 2000).

**DKEFS Inhibition:** This is a 3-part test of cognitive flexibility, ability to suppress a habitual response in favor of an unusual one, and selective attention. First, the examinee names the colors of boxes (red, green, or blue). Next, the examinee reads the names of the colors (“red,” “green,” “blue”) that appear in black print. Finally, the examinee responds to the



interference card in which the words are printed in non-corresponding color (e.g., “red” printed in blue ink) with the instructions to ignore the printed words and report only the color of ink in which the word is printed. For this study we examined the color-word inhibition score which provides a measure of susceptibility to interference; higher scores indicate better performance.

***Paced Auditory Serial Addition Test (PASAT) (Gronwall, 1977):*** The PASAT is a measure of cognitive function that specifically assesses auditory information processing speed and flexibility, attention, and calculation ability. Single digits are presented every two seconds and the participant must add each new digit to the one immediately prior to it. Scores reflect the number of correct sums given (maximum 50).

***Peabody Picture Vocabulary Test, IV Edition (PPVT-IV) (Dunn & Dunn, 2007):*** This is a measure of expressive vocabulary and word retrieval in which participants are asked to identify which of four pictures reflects a specific word. Scores are based on the number correct (maximum 228).

***The NIH Toolbox Dimensional Change Card Sorting (DCCS) Test (Zelazo et al., 2013):*** The Dimensional Change Card Sort Test is designed to assess the set shifting component of executive functioning. A target visual stimulus is displayed and the participant must match this to 1 of 2 choice stimuli according to shape or color. A criterion word “shape” or “color” appears at the bottom of the screen. For adults, the scoring algorithm is based on both accuracy and reaction time.

***The NIH Toolbox Flanker Inhibitory Control Test (Zelazo et al., 2013):*** The “Flanker” test measures the ability of the participant to inhibit responses to irrelevant task dimensions. On each item, a target stimulus (fish or arrow) is presented in the center surrounded on both sides by similar stimuli pointing in the same or different direction as the target stimuli. The examinee must indicate the direction that the central target stimulus is pointing and inhibit a response based upon the “flanking” stimuli. The scoring algorithm integrates accuracy and reaction time.

***Wisconsin Card Sorting Test-64 Card Version (Kongs, Thompson, Iverson, & Heaton, 2000)*** is a shortened, 64-card version of the Wisconsin Card Sorting Test, which assesses the ability to shift sets using visual stimuli that are easily verbally mediated. It requires participants to sort pictured cards into piles according to changing rules. As the rules for sorting change, the examinee is required to mentally shift sets and the test requires cognitive flexibility. The perseverative errors score was used as a criterion measure for this study.

**Data Analysis:** This study examines associations of the Toolbox List Sorting Working Memory Test scores with age, its test-retest reliability, and its convergent and discriminant validity. Age associations reflect the validity of the Toolbox Cognition tests for detecting age-related cognitive decline during adulthood. Test-retest correlations provide information on the reliability and stability of the measure. Convergent and discriminant validity results provide evidence that the Cognition Battery is measuring the intended constructs.

For the Toolbox and “gold standard” tests, raw scores for each subtest were converted to normally distributed scaled scores that were not age-corrected. These standard scores were created by first ranking the raw scores and then applying a normative transformation to the ranks to create a standard normal distribution, and as a final step, rescaling the distribution to have a mean of 10 and a standard deviation of 3. These scaled scores were used for all analyses and are not adjusted for age or any other demographic variable. Pearson correlation coefficients between actual age and test performance were calculated to assess the ability of the Toolbox List Sorting Working Memory Test to detect age-related cognitive decline during adulthood. Intraclass correlation coefficients (ICC) with 95% confidence intervals were calculated to evaluate test-retest reliability. ICCs are more likely than Pearson correlation coefficients to be affected by practice effects or other substantial differences that occur during the re-test interval. As a result, Pearson correlation coefficients tend to overestimate reliability coefficients. We computed practice effects, or the change in test scores between the test-retest administrations, by subtracting the performance at time 1 from the performance at time 2. We calculated the standardized estimate of the mean change (or “effect size”) by dividing the change score by the standard deviation of the time 1 score (Cohen, 1992). Convergent validity was assessed with correlations between the List Sorting Task and well-established “gold standard” working memory measures (e.g., WAIS-IV Letter-Number Sequencing, PASAT). Also, because there has been shown to be a relation between working memory and executive functioning tests, convergent validity was demonstrated between the List Sorting Task and measures of executive function (e.g., D-KEFS Color word task, Wisconsin Card Sorting Test, DCCS, Flanker). Finally, evidence of discriminant validity consisted of lower correlations with a “gold standard” measure of a *different* cognitive construct (i.e., receptive memory as assessed by the PPVT-4). To examine the effect of other demographic associations with performance, comparisons were performed using general linear models adjusted for age, education, and gender. Effect sizes are reported as Cohen’s *d*, with cutoffs of .20, .50, and .80 indicating small, medium, and large effects, respectively.

## Results

Of the 268 participants in the study, 264 completed the NIHTB List Sorting Test. The minimum score obtained in the sample was 8 points and the maximum score was 28 points; however, only 1 individual in the sample received a score at the ceiling of 28 points. The median score was 18 while the mean score was 18.32 with a standard deviation of 2.99 and a median score of 18.

### Age Effects

Cognitive abilities are expected to peak in early adulthood and then gradually decline with age. Participants ages 20–60 had average scaled scores of 10.84 (SD=2.78) whereas participants ages 65–85 had average scores of 8.75 (SD=2.80) (see Figure 2). A negative association was seen between age and List Sorting test performance ( $df = 264$ ,  $r = -0.44$ ,  $p < .0001$ ); successive age groups tended to complete fewer items correctly during middle and later adulthood.

## Other Demographic Associations

Table 1 shows effect sizes for other demographic factors known to influence cognitive performance, after adjusting for age and other additional relevant demographic variables. There was no significant difference in List Sorting Test or gold standard scores between males and females, with negligible to small effect sizes of 0.00 to 0.15. Caucasian participants scored better than African American or Hispanic participants on all measures, with moderate effect sizes between 0.41 and 0.58. Significant differences were also demonstrated for education, general health, and school problems for the List Sorting Test and gold standard measures.

## Test-Retest Reliability

For the participants who completed the retest 7 to 21 days after the initial test, test-retest ICCs were computed and demonstrated a high degree of consistency. Overall ICC for the Toolbox List Sorting Working Memory Test was .77 (95% CI = .67, .84) which was parallel to the test-retest reliability ICC for the criterion measure, the WAIS-IV Letter Number Sequencing, ( $r=0.80$ ; 95% CI = .71, .86). The List Sorting Test exhibited a significant by small practice effect. The mean practice effect in scaled score units = 0.79, SD = 1.89,  $t(88) = 3.95$ ,  $p = .001$ ,  $d=0.27$ . This is very similar to the practice effect obtained on the WAIS-IV Letter Number sequencing (mean= 0.49; SD = 1.76;  $t(88) = 2.59$ ;  $p = .01$ ;  $d = 0.18$ ).

## Convergent and Discriminant Validity

Table 2 shows results for convergent and discriminant validity. Convergent validity was demonstrated through the correlations between the List Sorting test and the WAIS-IV Letter Number Sequence ( $r=0.57$ ,  $p < .0001$ ), the processing speed measures ( $r = .50$  to  $r = .52$ ; all  $p < .0001$ ), and the measures of executive function ( $r = .43$  to  $r = .58$ ; all  $p < .0001$ ). We had predicted that this correlation (between two WM tasks) and between WM and executive functioning tests would be larger than the correlations between the List Sorting task and a truly divergent task, namely receptive vocabulary. Hence, discriminant validity was demonstrated through the correlations between the List Sorting test and the PPVT which had the lowest-magnitude correlation of  $r = .24$  ( $p < .0001$ ). Finally, Fisher's Exact Tests were conducted to statistically compare correlations between the List Sorting test and WAIS-III Working Memory versus the List Sorting Test and PPVT ( $r=0.57$  vs.  $r=0.24$ ;  $p < 0.001$ ).

## Discussion

The NIH Toolbox List Sorting Test is a new sequencing task that is designed to measure WM in 7 – 85 year olds. The task contains series of stimuli that have to be sequenced in size order. In the initial part of the test, the examinee sequences one type of stimulus (e.g., animals) in size order, from smallest to largest). The examinee is then asked to sequence stimuli from two different categories. The second part of the test, a 2-list task, requires examinees to both “bin” the stimuli into one of two categories (animals or food) while simultaneously sequencing each subgroup of stimuli in size order, from smallest to largest.

Successful WM tests tax the limit of an individual's storage capacity with a goal of determining when the WM system becomes overloaded. WM tasks often involve multi-

tasking activities such as reading a series of sentences and remembering the last word, reciting numbers while performing arithmetic tasks, or remembering a string of numbers and reciting what number occurred two or three numbers back. The two component tasks are often quite different and as such, these activities can be frustrating to the examinee. A strength of the List Sorting Test is that it progressively taxes WM capacity while remaining simple and straightforward. The test can be completed by children older than 7 years of age (Tulsky et al., 2013) and by adults through at least age 85.

Unlike traditional tests of WM that focus on either verbal functioning (PL) or visual stimuli (VS), the NIH Toolbox List Sorting Test presents each stimulus in both visual and auditory modalities. The List Sorting Test was designed to provide a general assessment of WM ability rather than provide specific, targeted assessment of the verbal or visual structures underlying WM ability. For a measurement tool like the NIH Toolbox, where the objective is to provide a rapid assessment of WM ability that can be repeated over time, the data presented here support such a claim.

The age trend data in this study reflect the pattern of gradual decline through adulthood that one would expect from a test of this kind (Dempster, 1981; Gathercole, Pickering, Ambridge, & Wearing, 2004; Heaton, Taylor, & Manly, 2003; McAuley & White, 2011; Wechsler, 1952). Performance tends to increase, as expected, through the childhood years, peaking in the early 20s. Test performance then tends to decrease through adulthood. The performance pattern by age mirrors the pattern on other measures of WM as well as other performance based tests.

Effects of additional demographic factors on the Toolbox List Sorting Working Memory Test were generally small to moderate, and consistent in magnitude with corresponding gold standard measures. Results from the Toolbox List Sorting Test parallel results on other, related cognitive tests (Heaton et al., 2003; Strauss, Sherman, & Spreen, 2006). Medium effect sizes were found for race/ethnicity and education, a finding that is typical of other WM measures. Furthermore, with regard to ecological validity, the List Sorting Test has small to moderate relationships with current self-reported general health status as well as self-reported history of problems in school.

As predicted, the List Sorting Test demonstrated test-retest reliability with a test-retest correlation coefficient of .77. This is comparable to other established tests of WM such as the WAIS-IV Letter Number Sequencing task where the coefficient was .80. Furthermore, study data provide evidence of convergent and discriminant validity of the List Sorting Test as a measure of WM and executive function and also provide initial support for the construct validity of the List Sorting Test. Convergent validity is seen in the correlation between the List Sorting Test and other measures of WM (i.e., WAIS-IV Letter Number Sequencing and the PASAT). The correlations of  $r = .57$  and  $r = .50$ , respectively, provide evidence that both tests are measuring a similar construct. Similarly, discriminant validity is seen in the relatively low correlation between the List Sorting Test and the PPVT, a test of receptive vocabulary ( $r = .24$ ).

It is important to note that while the List Sorting Test is related to criterion measures of WM, it is also moderately correlated with tests measuring executive functioning including tests of inhibition, cognitive flexibility, perseveration, and processing speed. This is not a unique finding as several previous investigators have reported high correlations between WM and executive functioning (Lehto, 1996; Miyake et al., 2000) and other measures of general reasoning (Kyllonen & Christal, 1990). The List Sorting Test had moderately high correlations with the WAIS-IV Processing Speed Index ( $r = .52$ ) and also with tests of executive functioning (i.e., WCST, DKEFS, DCCS, and Flanker; correlations ranging from  $r = .43$  to  $r = .58$ ). While not uncommon in a test of WM, examiners should be aware that the List Sorting Test is related to these additional constructs. Further research will help the research and clinical communities determine if the List Sorting Test is a specific measure of WM or a more global measure of executive functioning.

The validation of a new test is always a process. There is not one study or one statistic that captures construct validity. Originally introduced by Cronbach and Meehl (1955), the concept of construct validity was updated by Samuel Messick (1980) to entail a process of “marshaling” data from several experiments where the researcher forms hypothesis and uses empirical data to confirm that the test performs as expected. Messick (1991) then went a step further to assert that all types of validity (e.g. content, criterion, discriminant) were really subcomponents of construct validation. The initial data presented here provide the initial empirical support of the NIH Toolbox List Sorting Test as a valid measure of WM. These data can be “marshaled together” with the results presented previously by Tulsky et al. (2013) that showed that List Sorting Test’s performance in children also met the a priori research hypotheses.

At the same time, the results in this manuscript stop short of providing the conclusive validity evidence necessary for utilization of the NIH Toolbox List Sorting Test in clinical populations for either clinical trials research or clinical uses. Furthermore, it has yet to be determined if the List Sorting task will measure WM independent of general executive functioning. These limitations should not be regarded as negative properties of the List Sorting Test, but rather as a limitation of the data that has been collected to date and a caution until further research is performed. Additional studies are needed to evaluate the performance of this test (and the other NIH Toolbox component tests) in clinical groups. That withstanding, the results presented here do provide support for the validity and reliability of the NIH Toolbox List Sorting Test.

## Acknowledgments

Funding for this research was provided by contract number N01AG60007 from the National Institutes of Health.

## References

- Alloway TP, Gathercole SE, Pickering SJ. Verbal and visuospatial short-term and working memory in children: are they separable? *Child Development*. 2006; 77(6):1698–1716. [PubMed: 17107455]
- Awh E, Jonides J, Smith EE, Schumacher EH, Koeppel RA, Katz S. Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychol Sci*. 1996; 7(1):25–31.
- Baddeley, A. *Working Memory*. Oxford, England: Oxford University Press; 1986.

- Baddeley, A. Working Memory. Vol. 11. Clarendon Press; 1987.
- Baddeley A. Working memory. *Science*. 1992; 255(5044):556–559. [PubMed: 1736359]
- Baddeley A. Exploring the central executive. *Quarterly Journal of Experimental Psychology*. 1996; 49A:5–28.
- Baddeley A. The episodic buffer: a new component of working memory? *Trends Cogn Sci*. 2000; 4(11):417–423. [PubMed: 11058819]
- Baddeley A. Is Working Memory Still Working? *American Psychologist*. 2001; 56(11):851–864. [PubMed: 11785152]
- Baddeley A. Is working memory still working? *European Psychologist*. 2002; 7(2):85–97.
- Baddeley A. Working Memory. *Current Biology*. 2010; 20(4):R136–R140. [PubMed: 20178752]
- Baddeley, A.; Hitch, GJ. Working Memory. In: Bower, GH., editor. *The psychology of learning and motivation: Advances in research and theory*. Vol. 8. San Diego, CA: Academic Press; 1974. p. 47-90.
- Beatty WW, Wilbanks SL, Blanco CR, Hames KA, Tivis R, Paul RH. Memory disturbance in multiple sclerosis: reconsideration of patterns of performance on the selective reminding test. *Journal of Clinical and Experimental Neuropsychology*. 1996; 18(1):56–62. 10.1080/01688639608408262 [PubMed: 8926297]
- Belger A, Puce A, Krystal JH, Gore JC, Goldman-Rakic P, McCarthy G. Dissociation of mnemonic and perceptual processes during spatial and nonspatial working memory using fMRI. *Human Brain Mapping*. 1998; 6(1):14–32. [PubMed: 9673660]
- Braver TS, Cohen JD, Nystrom LE, Jonides J, Smith EE, Noll DC. A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage*. 1997; 5:49–62. [PubMed: 9038284]
- Chiaravalloti N, Hillary F, Ricker J, Christodoulou C, Kalnin A, Liu WC, DeLuca J. Cerebral activation patterns during working memory performance in multiple sclerosis using FMRI. *Journal of Clinical and Experimental Neuropsychology*. 2005; 27(1):33–54. [PubMed: 15814441]
- Cohen J. A power primer. *Psychological Bulletin*. 1992; 112:155–159. [PubMed: 19565683]
- Conlin JA, Gathercole SE, Adams JW. Children’s working memory: investigating performance limitations in complex span tasks. *Journal of Experimental Child Psychology*. 2005; 90(4):303–317. [PubMed: 15777923]
- Conway AR, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, Engle RW. Working memory span tasks: A methodological review and user’s guide. *Psychonomic Bulletin & Review*. 2005; 12(5): 769–786. [PubMed: 16523997]
- Courtney SM, Ungerleider LG, Keil K, Haxby JV. Transient and sustained activity in a distributed neural system for human working memory. *Nature*. 1997; 386(6625):608–611. [PubMed: 9121584]
- Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychological Bulletin*. 1955; 52:281–302. [PubMed: 13245896]
- Crosson B, Rao SM, Woodley SJ, Rosen AC, Bobholz JA, Mayer A, Stein EA. Mapping of semantic, phonological, and orthographic verbal working memory in normal adults with functional magnetic resonance imaging. *Neuropsychology*. 1999; 13(2):171–187. [PubMed: 10353369]
- Crowe SF. Does the letter number sequencing task measure anything more than digit span? *Assessment*. 2000; 7(2):113–117. [PubMed: 10868248]
- Curtis CE, D’Esposito M. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn Sci*. 2003; 7(9):415–423. [PubMed: 12963473]
- D’Esposito M, Aguirre GK, Zarahn E, Ballard D, Shin RK, Lease J. Functional MRI studies of spatial and nonspatial working memory. *Cognitive Brain Research*. 1998; 7(1):1–13. [PubMed: 9714705]
- D’Esposito M, Onishi K, Thompson H, Robinson K, Armstrong C, Grossman M. 1996; 10:51–56. W. m. i. i. m. s. N. Working memory impairments in multiple sclerosis. *Neuropsychology*. 1996; 10:51–56.
- de Jong PF. Working memory deficits of reading disabled children. *Journal of Experimental Child Psychology*. 1998; 70(2):75–96. [PubMed: 9729450]
- de Jong PF, Olson RK. Early predictors of letter knowledge. *Journal of Experimental Child Psychology*. 2004; 88(3):254–273. [PubMed: 15203300]



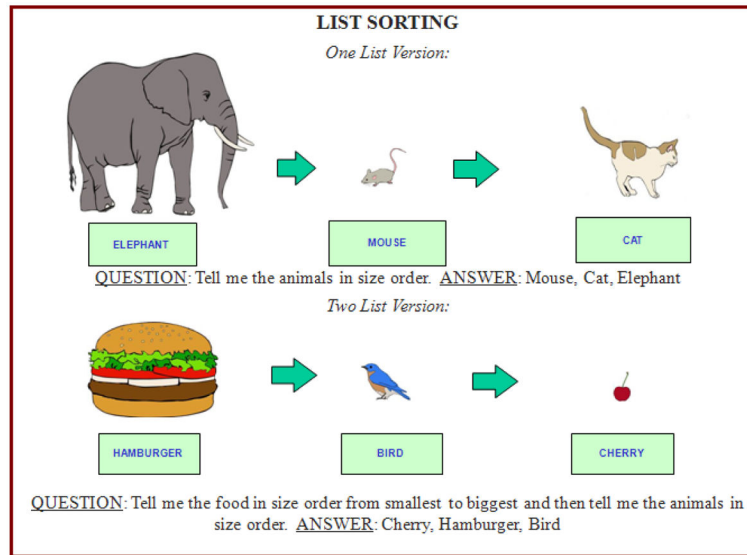
- Demaree HA, DeLuca J, Gaudino EA, Diamond BJ. Speed of information processing as a key deficit in multiple sclerosis: implications for rehabilitation. *Journal of Neurology, Neurosurgery, & Psychiatry*. 1999; 67(5):661–663.
- Dempster FN. Memory Span - Sources of Individual and Developmental Differences. *Psychological Bulletin*. 1981; 89(1):63–100.
- Donders J, Tulsky DS, Zhu J. Criterion validity of new WAIS-II subtest scores after traumatic brain injury. *Journal of the International Neuropsychological Society*. 2001; 7(7):892–898. [PubMed: 11771633]
- Dunn, LM.; Dunn, DM. Peabody Picture Vocabulary Test. 4. Minneapolis, MN: NCS Pearson; 2007.
- Edin F, Klingberg T, Johansson P, McNab F, Tegner J, Compte A. Mechanism for top-down control of working memory capacity. *Proc Natl Acad Sci U S A*. 2009; 106(16):6802–6807. [PubMed: 19339493]
- Edin F, Macoveanu J, Olesen P, Tegner J, Klingberg T. Stronger synaptic connectivity as a mechanism behind development of working memory-related brain activity during childhood. *Journal of Cognitive Neuroscience*. 2007; 19(5):750–760. [PubMed: 17488202]
- Engle, RW. Working memory and retrieval: An inhibition-resource approach. In: Richardson, J., editor. *Working Memory and Human Cognition*. New York: Oxford University Press; 1996. p. 89-119.
- Fuster, JM. *The Prefrontal Cortex*. New York: Raven Press; 1989.
- Gathercole SE, Pickering SJ, Ambridge B, Wearing H. The structure of working memory from 4 to 15 years of age. *Developmental Psychology*. 2004; 40(2):177–190. [PubMed: 14979759]
- Gershon RC, Cella D, Fox NA, Havlik RJ, Hendrie HC, Wagster MV. Assessment of neurological and behavioural function: the NIH Toolbox. *Lancet Neurology*. 2010; 9(2):138–139.
- Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, Nowinski CJ. NIH Toolbox for Assessment of Neurological and Behavioral Function. *Neurology*. 2013; 80(Suppl 3):S2–S6. [PubMed: 23479538]
- Gold JM, Carpenter C, Randolph C, Goldberg TE, Weinberger DR. Auditory working memory and Wisconsin Card Sorting Test performance in schizophrenia. *Archives of General Psychiatry*. 1997; 54(2):159–165. [PubMed: 9040284]
- Goldman-Rakic, PS. Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: Mountcastle, VB., editor. *Handbook of Physiology*. Bethesda, MD: American Psychological Society; 1987. p. 373-417.
- Grigsby J, Ayarbe S, Kravcisin N, Busenbark D. Working memory impairment among persons with chronic-progressive multiple sclerosis. *Journal of Neurology*. 1994; 241(3):125–131. [PubMed: 8164013]
- Grigsby J, Busenbark D, Kravcisin N, Kennedy PM, Taylor D. Impairment of the working memory system in relapsing-remitting multiple sclerosis. *Archives of Clinical Neuropsychology*. 1994; 9:134–135.
- Gronwall DM. Paced auditory serial-addition task: a measure of recovery from concussion. *Perceptual & Motor Skills*. 1977; 44(2):367–373. [PubMed: 866038]
- Grossman M, Armstrong C, Onishi K, Thompson H, Schaefer B, Robinson K, Silberberg D. Patterns of Cognitive Impairment in Relapsing-Remitting and Chronic Progressive Multiple-Sclerosis. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*. 1994; 7(3):194–210.
- Hasher L, Stolzhus ER, Zacks RT, Rypma B. Age and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1991; 17(1):163–169.
- Haut MW, Kuwabara H, Leach S, Arias RG. Neural activation during performance of number-letter sequencing. *Applied Neuropsychology*. 2000; 7(4):237–242. [PubMed: 11296686]
- Hawkins KA. Indicators of brain dysfunction derived from graphic representations of the WAIS-III/WMS-III Technical Manual clinical samples data: A preliminary approach to clinical utility. *The Clinical Neuropsychologist*. 1998; 12(4):535–551.
- Heaton, RK.; Taylor, MJ.; Manly, J. Demographic Effects and Use of Demographically Corrected Norms with the WAIS-III and WMS-III. In: Tulsky, DS., et al., editors. *Clinical Interpretation of the WAIS-III and WMS-III*. San Diego: Academic Press; 2003. p. 181-210.



- Hitch GJ, Towse JN, Hutton U. What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology-General*. 2001; 130(2):184–198. [PubMed: 11409098]
- Hodes RJ, Insel TR, Landis SC. The NIH Toolbox: Setting a Standard for Biomedical Research. *Neurology*. 2013; 80(Suppl 3):S1. [PubMed: 23479536]
- Hofmann W, Schmeichel BJ, Baddeley AD. Executive functions and self-regulation. *Trends Cogn Sci*. 2012; 16(3):174–180.10.1016/j.tics.2012.01.006 [PubMed: 22336729]
- Jacobsen CF. Functions of frontal association area in primates. *Archives of Neurology and Psychiatry*. 1935; 33(3):558–569.
- Jonides, J. Working memory and thinking. In: EES; DNO, editors. *Invitation to Cognitive Science: Thinking*. 2. Vol. 3. Cambridge, MA: MIT Press; 1995. p. 215-265.
- Kirasic K, Allen G, Dobson S, Binder K. Aging, cognitive resources, and declarative learning. *Psychology and Aging*. 1996; 11:658–670. [PubMed: 9000297]
- Klingberg T, Kawashima R, Roland PE. Activation of multi-modal cortical areas underlies short-term memory. *European Journal of Neuroscience*. 1996; 8(9):1965–1971. [PubMed: 8921287]
- Kongs, SK.; Thompson, LL.; Iverson, GL.; Heaton, R. *Wisconsin card sorting test - 64 card version: Professional Manual*. Odessa, FL: Psychological Assessment Resources; 2000.
- Kyllonen PC, Christal RE. Reasoning ability is (little more than) working-memory capacity? *Intelligence*. 1990; 14:389–433.
- Lehto J. Are executive function tests dependent on working memory capacity? *The Quarterly Journal of Experimental Psychology*. 1996; 49A(1):29–50.
- Linden DE. The working memory networks of the human brain. *Neuroscientist*. 2007; 13(3):257–267. [PubMed: 17519368]
- Litvan I, Grafman J, Vendrell P, Martinez JM. Slowed information processing in multiple sclerosis. *Archives of Neurology*. 1988; 45(3):281–285. [PubMed: 3341952]
- Logie, RH. The seven ages of working memory. In: Richardson, J., editor. *Working memory and human cognition*. New York: Oxford University Press; 1996. p. 31-65.
- Manoach DS, Schlag G, Siewert B, Darby DG, Bly BM, Benfield A, Warach S. Prefrontal cortex fMRI signal changes are correlated with working memory load. *Neuroreport*. 1997; 8(2):545–549. [PubMed: 9080445]
- Martin TA, Donders J, Thompson E. Potential of and problems with new measures of psychometric intelligence after traumatic brain injury. *Rehabil Psychol*. 2000; (45):402–408.
- McAllister TW, Flashman LA, Sparling MB, Saykin AJ. Working memory deficits after traumatic brain injury: catecholaminergic mechanisms and prospects for treatment -- a review. *Brain Injury*. 2004; 18(4):331–350. [PubMed: 14742148]
- McAllister TW, Saykin AJ, Flashman LA, Sparling MB, Johnson SC, Guerin SJ, Yanofsky N. Brain activation during working memory 1 month after mild traumatic brain injury: a functional MRI study. *Neurology*. 1999; 53(6):1300–1308. [PubMed: 10522888]
- McAuley T, White DA. A latent variables examination of processing speed, response inhibition, and working memory during typical development. *Journal of Experimental Child Psychology*. 2011; 108(3):453–468. [PubMed: 20888572]
- McCabe DP, Roediger HL, McDaniel MA, Balota DA, Hambrick DZ. The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*. 2010; 24:222–243. [PubMed: 20230116]
- Messick S. Test Validity and the Ethics of Assessment. *American Psychologist*. 1980; 35(11):1012–1027.
- Messick, S. Validity of Test Interpretation and Use. In: Alkin, MC., editor. *Encyclopedia of Educational Research*. 6. New York: Macmillan; 1991.
- Miller GA. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review*. 1956; 63(2):81–97. [PubMed: 13310704]
- Milner, B. Some effects of frontal lobectomy in man. In: Warren, JM.; Akert, K., editors. *The Frontal Granular Cortex and Behavior*. New York: McGraw-Hill; 1964. p. 313-334.

- Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*. 2000; 41:49–100. [PubMed: 10945922]
- Miyake, A.; Shah, P., editors. *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. New York: Cambridge University Press; 1999.
- Mungas D, Reed BR, Marshall SC, Gonzalez HM. Development of psychometrically matched English and Spanish language neuropsychological tests for older persons. *Neuropsychology*. 2000; 14(2): 209–223. [PubMed: 10791861]
- Newell, A. Productions systems: Models of control structures. In: Chase, WG., editor. *Visual Information Processing*. New York: Academic Press; 1973. p. 463-526.
- Newell, A.; Simon, HA. *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall; 1972.
- Orellana G, Slachevsky A. Executive functioning in schizophrenia. *Front Psychiatry*. 2013; 4:35.10.3389/fpsy.2013.00035 [PubMed: 23805107]
- Owen AM, Evans AC, Petrides M. Evidence for a two-stage model of spatial working memory processing within the lateral frontal cortex: a positron emission tomography study. *Cerebral Cortex*. 1996; 6(1):31–38. [PubMed: 8670636]
- Paulesu E, Frith CD, Frackowiak RS. The neural correlates of the verbal component of working memory. *Nature*. 1993; 362(6418):342–345. [PubMed: 8455719]
- Pelphrey KA, Reznick JS. Working memory in infancy. *Advances in Child Development and Behavior*. 2003; 31:173–227. [PubMed: 14528662]
- Rao SM, Grafman J, DiGiulio D, Mittenberg W, Bernardin L, Leo GJ, Unverzagt F. Memory dysfunction in multiple sclerosis: Its relation to working memory, semantic encoding and implicit learning. *Neuropsychology*. 1993; 7(3):364–374.
- Repovs G, Bresjanak M. Cognitive neuroscience of working memory. *Neuroscience*. 2006; 139:1–413.
- Riggs K, McTaggart J, Simpson A. Changes in the Capacity of Visual Working Memory in 5-to-10 - Year-Olds. *Journal of Experimental Child Psychology*. 2006; 95:18–26. [PubMed: 16678845]
- Salmon E, Van der Linden M, Collette F, Delfiore G, Maquet P, Degueldre C, Franck G. Regional brain activity during working memory tasks. *Brain*. 1996; 119(Pt 5):1617–1625. [PubMed: 8931584]
- Salthouse T. The Aging of Working Memory. *Neuropsychology*. 1994; 8(4):535–543.
- Salthouse T, Mitchell DR, Skovronek E, Babcock RL. Effects of adult age and working memory on reasoning and spatial abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1989; 15(3):507–516.
- Schmiedek F, Hildebrandt A, Lovden M, Lindenberger U, Wilhelm O. Complex span versus updating tasks of working memory: the gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2009; 35(4):1089–1096.
- Seidman LJ, Breiter HC, Goodman JM, Goldstein JM, Woodruff PWR, Rosen BR. A functional magnetic resonance imaging study of auditory vigilance with low and high information processing demands. *Neuropsychology*. 1998; 12(4):505–518. [PubMed: 9805320]
- Smith EE, Jonides J, Koeppel RA. Dissociating verbal and spatial working memory using PET. *Cerebral Cortex*. 1996; 6(1):11–20. [PubMed: 8670634]
- Strauss, E.; Sherman, EMS.; Spreen, O. *A compendium of neuropsychological tests: Administration, norms, and commentary*. 3. Oxford; New York: Oxford University Press; 2006.
- Tulsky DS, Carlozzi NE, Chevalier N, Espy K, Beaumont J, Mungas D. NIH Toolbox Cognitive Function Battery (CFB): Measuring Working Memory. *Society For Research In Child Development, Monograph*. 2013; 78(4):70–87.
- Tulsky, DS.; Saklofske, DH.; Zhu, J. *Clinical Interpretation of the WAIS-III and WMS-III*. San Diego: Elsevier Science; 2003. Revising a Standard: Evaluation of the Origin and Development of the WAIS-III; p. 43-92.
- Van Snellenberg JX. Working memory and long-term memory deficits in schizophrenia: is there a common substrate? *Psychiatry Research*. 2009; 174(2):89–96. [PubMed: 19837568]
- Vugs B, Hendriks M, Cuperus J, Verhoeven L. Working memory performance and executive function behaviors in young children with SLI. *Res Dev Disabil*. 2014; 35(1):62–74. [PubMed: 24240018]

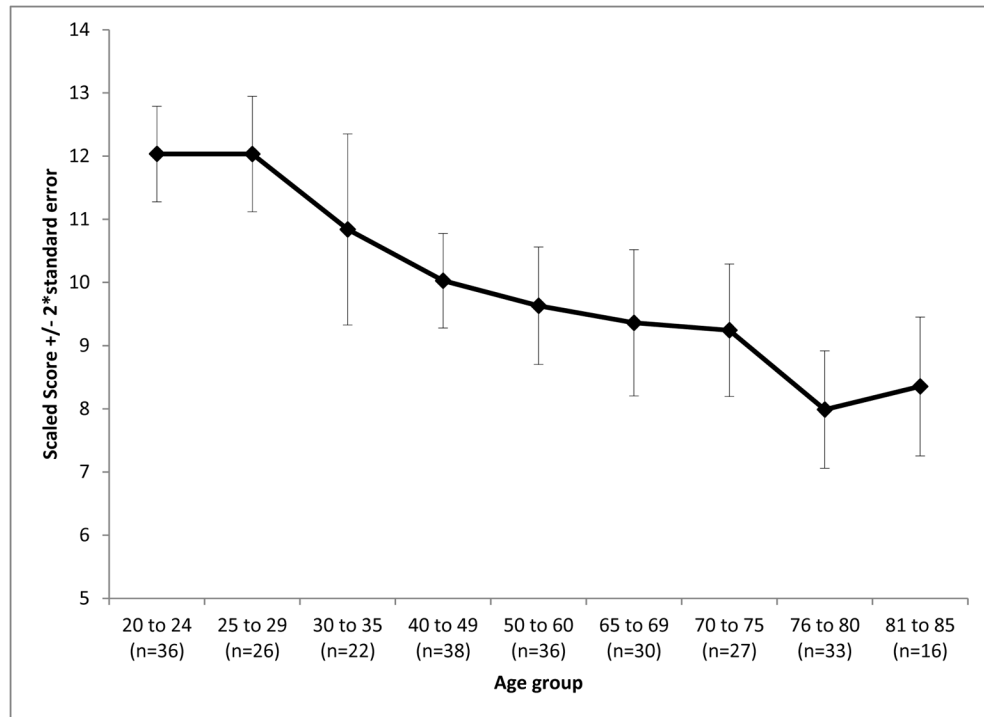
- Wechsler, D. *The Range of Human Capacities*. Baltimore: The Williams & Wilkins Company; 1952.
- Wechsler, D. *WAIS-III Administration and Scoring Manual*. San Antonio: The Psychological Corporation; 1997.
- Wechsler, D. *Wechsler Adult Intelligence Scale IV*. San Antonio: Harcourt Assessment Inc; 2008.
- Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, Gershon RC. Cognition assessment using the NIH Toolbox. *Neurology*. 2013; 80(11 Suppl 3):S54–64. [PubMed: 23479546]
- Willmott C, Ponsford J, Hocking C, Schönberger M. Factors contributing to attentional impairments after traumatic brain injury. *Neuropsychology*. 2009; 23(4):424–432. [PubMed: 19586207]
- Woods SP, Moore DJ, Weber E, Grant I. Cognitive neuropsychology of HIV-associated neurocognitive disorders. *Neuropsychology Review*. 2009; 19(2):152–168. [PubMed: 19462243]
- You SC, Geschwind MD, Sha SJ, Apple A, Satris G, Wood KA, Possin KL. Executive functions in premanifest Huntington’s disease. *Movement Disorders*. 2013
- Zelazo PD, Andersen J, Richler J, Wallner-Allen K, Beaumont J, Weintraub S. *NIH Toolbox Cognitive Function Battery (CFB): Measuring Executive Function and Attention*. Society For Research In Child Development, Monograph. 2013; 78(4):16–33.



**Figure 1.**

Examples of One-List and Two-List List Sorting Task

1-List List Sorting requires participants to sequence items according to a single category, whereas 2-List List Sorting requires sequencing that involves an alternation between two different categories.



**Figure 2.** Toolbox List Sorting Working Memory Test scores by age (means  $\pm$  2 standard errors).

**Table 1**

Effect sizes (ES) for comparisons of scores between groups

	Toolbox List Sorting	Working Memory Test	WAIS-IV Letter-Number Sequencing	PASAT
ES (Male vs. Female) <sup>1</sup>	0.01		0.00	0.15
<i>P</i>	0.910		0.992	0.212
ES (Black vs. White) <sup>2</sup>	-0.54		-0.55	-0.48
ES (Hispanic vs. White) <sup>2</sup>	-0.41		-0.58	-0.50
<i>P</i>	<0.001		<0.001	<0.001
ES (College vs. < HS) <sup>3</sup>	0.35		0.52	0.59
ES (College vs HS Grad) <sup>3</sup>	-0.02		0.38	0.24
ES (HS Grad vs. < HS) <sup>3</sup>	0.37		0.14	0.35
<i>P</i>	0.006		0.003	<0.001
ES (Excellent or very good health vs. < good) <sup>3</sup>	0.40		0.39	0.52
ES (Excellent or very good health vs. < fair-poor) <sup>3</sup>	0.60		0.61	0.67
<i>P</i>	<0.001		<0.001	<0.001
ES (No school problems vs any) <sup>3</sup>	0.22		0.37	0.23
<i>P</i>	0.046		0.002	0.064

Note.

<sup>1</sup> Gender comparison adjusted for age and education;

<sup>2</sup> Race/ethnicity comparisons adjusted for gender, age, and education;

<sup>3</sup> Other comparisons adjusted for age

**Table 2**

Convergent &amp; Discriminant Validity of Toolbox List Sorting Working Memory Test

	<b>df</b>	<b>R</b>
<b>Convergent Validity Measures</b>		
WAIS-IV Letter-Number Sequencing	261	.57
PASAT	256	.50
D-KEFS Color-Word Interference	256	.58
WCST-64 Perseverative Errors	258	.54
WAIS-IV PS Composite	263	.52
DCCS	243	.44
Flanker	235	.43
<b>Discriminant Validity Measure</b>		
PPVT-4	262	.24

*Note.* All  $p < .0001$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript