Cognitive, Affective, & Behavioral Neuroscience
https://doi.org/10.3758/s13415-025-01310-3

RESEARCH ARTICLE

The latent structure of working memory: A large sample factor model of working memory capacity

Han Hao[1] · John C. Williams[2,3] · Philip N. Tubiolo[2,3] · Kacie Bauer[1] · Eilon Silver-Frankel[2,4] · Sam R. Luceno[2] · Avery J. Chan[2,5] · Zu Jie Zheng[2,6] · Kelly R. Bobchin[2] · Roman Kotov[2] · Greg Perlman[2] · Andrew R. A. Conway[7] · Jared X. Van Snellenberg[2,3,4]

Accepted: 2 May 2025

Abstract
Working memory (WM) is an essential system of cognitive processes for a wide range of cognitive activities and is associated with diverse real-world outcomes. Despite extensive research in cognitive psychology, the complex multifaceted nature of WM is often overlooked in applied settings, such as clinical and neuroimaging research. This study investigated the latent structure of WM by examining a comprehensive set of WM tasks commonly used in both theoretical and applied research in cognitive psychology and psychiatric neuroimaging. A large sample of healthy, young adults ($N = 608$) completed a battery of WM tasks and other cognitive measures. Factor analyses and structural equation models revealed a three-factor structure: Storage, Executive Attention, and Updating. These factors were moderately correlated but contributed uniquely to explaining variance in intelligence measures. Furthermore, when the three factors were considered in a single model, only the Updating and Executive Attention factors had unique shared variance with intelligence. The findings support that WM is a multifaceted construct, with complex span and n-back tasks capturing important and distinct components related to real-world cognitive performance. This highlights the need for precise selection of measurement tools for WM in both theoretical and applied research contexts.

Keywords Working memory · Cognitive individual differences · Latent factor analysis · Structural equation modeling

Introduction

Working memory (WM) is a system of cognitive processes that maintains temporary access to a limited amount of information for ongoing processing in complex cognition (Baddeley & Graham, 1974; Baddeley, 2010; Cowan, 2014). It is an essential system for a wide range of mental activities, including planning, reasoning, learning, and comprehension (Cowan et al., 2024). Individual differences in WM are related to cognition generally (Kovacs & Conway, 2016) and are predictive of real-world outcomes, including reading comprehension (Daneman & Carpenter, 1980), mathematics performance (Ramirez et al., 2013), and writing ability (Swanson & Berninger, 1996). Hence, it is crucial to understand the function and structure of WM across different fields of psychology and psychiatry. For example, previous work investigating cognitive abilities in people with schizophrenia (SCZ) indicates that WM deficits are a hallmark symptom (Lee & Park, 2005; Forbes et al., 2009) closely associated with poor functional outcomes (Goldman-Rakic,

John C. Williams and Philip N. Tubiolo contributed equally to this work.

✉ Han Hao
hhao@tarleton.edu

1 Department of Psychological Sciences, Tarleton State University, Stephenville, TX, USA

2 Department of Psychiatry and Behavioral Health, Stony Brook University Renaissance School of Medicine, Stony Brook, NY, USA

3 Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY, USA

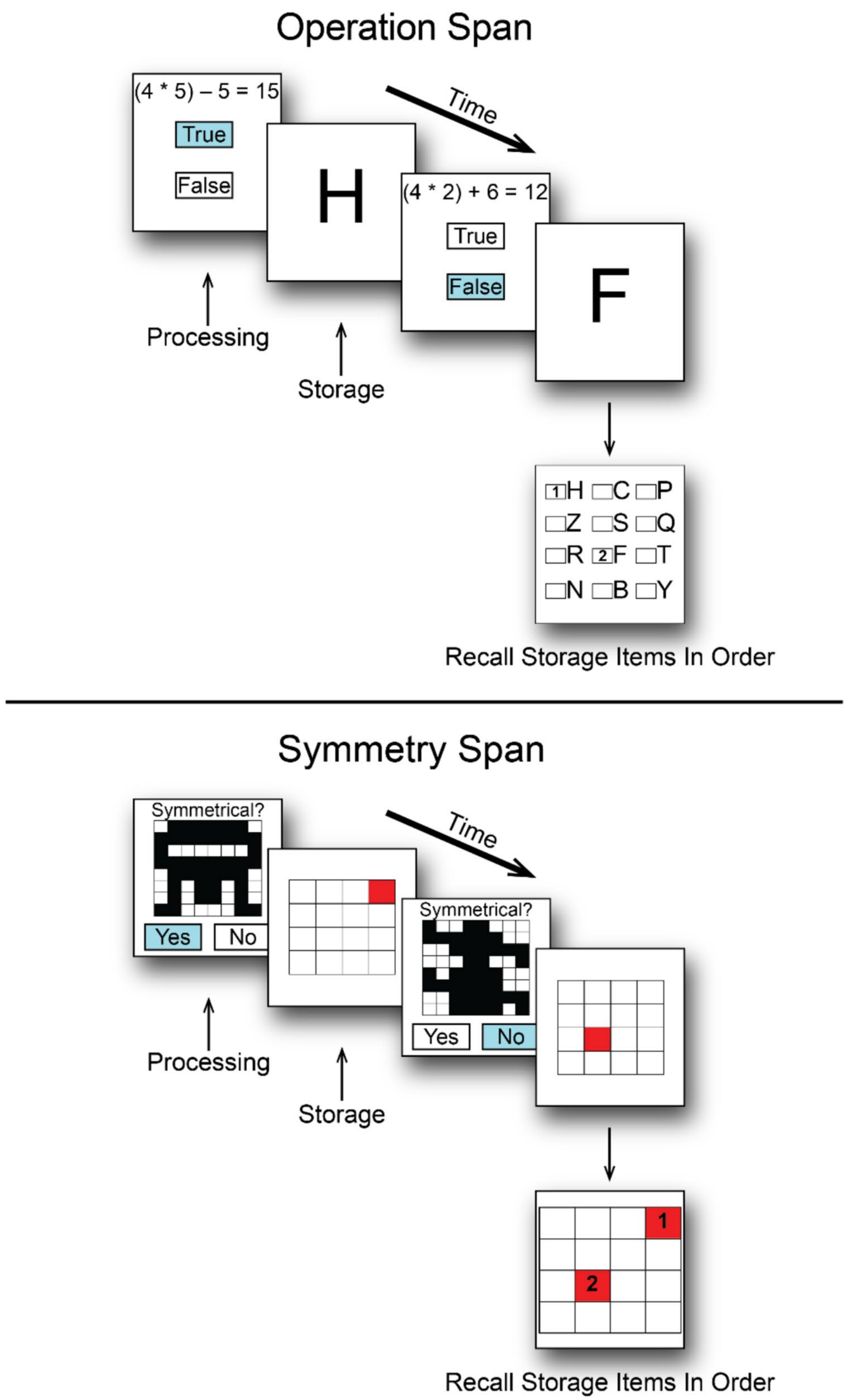
4 Department of Psychology, Stony Brook University, Stony Brook, NY, USA

5 Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY, USA

6 College of Medicine, State University of New York Downstate Health Sciences University, Brooklyn, NY, USA

7 Department of Psychology, New Mexico State University, Las Cruces, NM, USA

Published online: 02 June 2025



1994, 1999; Bowie & Harvey, 2006; Green, 1998; Green & Harvey, 2014; Kern et al., 2011). In addition, WM deficits are observed in a wide range of neuropsychiatric disorders (Abramovitch et al., 2021). However, despite extensive studies on WM in both basic and clinical/translational research, the construct validity of WM and the specific cognitive

◂**Fig. 1** **Top**: Schematic of an Operation Span task item. In each item there are multiple trials of stimuli, typically consisting of a processing component (a mathematical equation) and a storage component (a letter). On each trial participants must judge whether the mathematical equation is correct and memorize the letter. After a few trials, they need to recall the letters in their presented order. **Bottom**: Schematic of a Symmetry Span task item. In each item there are multiple trials of stimuli, typically consisting of a processing component (a figure of colored 8x8 grid) and a storage component (a random colored cell in a 4x4 grid). On each trial participants must judge whether the figure is symmetrical and then memorize the position of the colored cell in the 4x4 grid. After a few trials, they need to recall the cells in their presented order

mechanisms underlying performance of various WM tasks remain areas of active investigation.

Traditional psychological models of WM, such as Baddeley and Hitch's multicomponent model (Baddeley & Hitch, 1974), proposed distinctions among different components of WM, including domain-general central executive attention mechanisms and domain-specific memory storage systems. In contrast to Baddeley & Hitch, which emphasizes the *structure* of WM and different *stores* of memory, Cowan's (1999) embedded processes model of WM centers on the *function* of WM and different *states* of memory representations, positing that WM consists of the temporarily activated portion of long-term memory managed by a central executive through focused attention. This emphasis on the role of attention during information processing inspired more recent theoretical models of WM to pinpoint the complex and flexible mechanisms by which information is activated and maintained. For example, Oberauer & Hein (2012) distinguished three components in WM: activated long-term memory, the region of direct access, and a single-item focus of attention. The theory emphasized that WM is an attention selection system to memory representations with temporary bindings of objects/features (Oberauer, 2019a), and that individual differences in WM reflect individual differences in a specific limit on establishing and maintaining these bindings (Oberauer, 2019b).

Working memory measures in different research areas

Although they differ in their emphasis on structure vs. function, different theoretical models of WM generally agree that WM is a complex, multi-faceted system with limited capacity. Thus, various tasks have been developed to measure individual differences in working memory capacity (WMC). These tasks typically involve active maintenance of to-be-recalled information and simultaneous processing of other information. However, different research areas seem to have different preferences when it comes to specific WM task paradigms.

For example, in the cognitive psychology and individual difference literature, complex span tasks (Conway et al., 2005) are the most common tasks used to measure WMC. Unlike simple span tasks, a common measure of short-term memory, complex span tasks consist of a dual demand that requires participants to store and process information simultaneously. In this paradigm, WMC is the amount of information that can be maintained in the face of concurrent processing. For example, in the operation span task, participants solve math problems (processing component) while remembering a series of numbers (storage component). Importantly, the storage and processing stimuli are from the same domain in any given complex span task, such that the processing task creates an intra-domain distraction condition that interferes with maintenance and rehearsal of the target information. For instance, in verbal/numerical complex span tasks, both the storage and processing stimuli are verbal/numerical, while in spatial complex span tasks, both the storage and processing stimuli are spatial. Measures of WMC from complex span tasks have demonstrated high reliability (Redick et al., 2012) and strong predictive validity (Conway et al., 2003; Schmiedek et al., 2009), making them widely applicable in cognitive psychology. Performance on complex span tasks correlates strongly with performance in experimental and real-world tasks that are theoretically assumed to require WM, such as reasoning (Kane et al., 2004) and language comprehension (Daneman & Merikle, 1996). Figure 1 shows the schematics of two commonly used complex span tasks.

In contrast, in the field of cognitive neuroscience, where brain imaging and electrophysiology studies are prevalent, n-back tasks (Owen et al., 2005) are most commonly used to measure WMC. Throughout the WM literature, there are two widely used variants of n-back tasks (Tubiolo et al., 2024). In one variant, herein referred to as the delayed-match-to-sample n-back (NB-DMS), participants are presented with a series of stimuli—such as letters, numbers, figures, or pictures—and make a target/nontarget judgment as to whether the current stimulus matches the one presented n items back (Moore & Ross, 1963; Ross, 1966a, b). The value of n can vary across experimental conditions to vary the memory and processing demands of the task, which is often referred to as WM load (Chen et al., 2008). The other variant of the n-back paradigm, referred to as the continuous delayed response n-back (NB-CDR), requires participants to respond to the current stimulus based on the stimulus presented n items back (with typically four response options that are one-to-one mapped with four stimuli), effectively delaying their response for n items (Callicott et al., 1998, 1999, 2000). Figure 2 shows the schematics of the two types of n-back tasks.

Despite the common application of n-back tasks in cognitive neuroscience, research has shown that performance on NB-DMS tasks and complex span tasks is only weakly

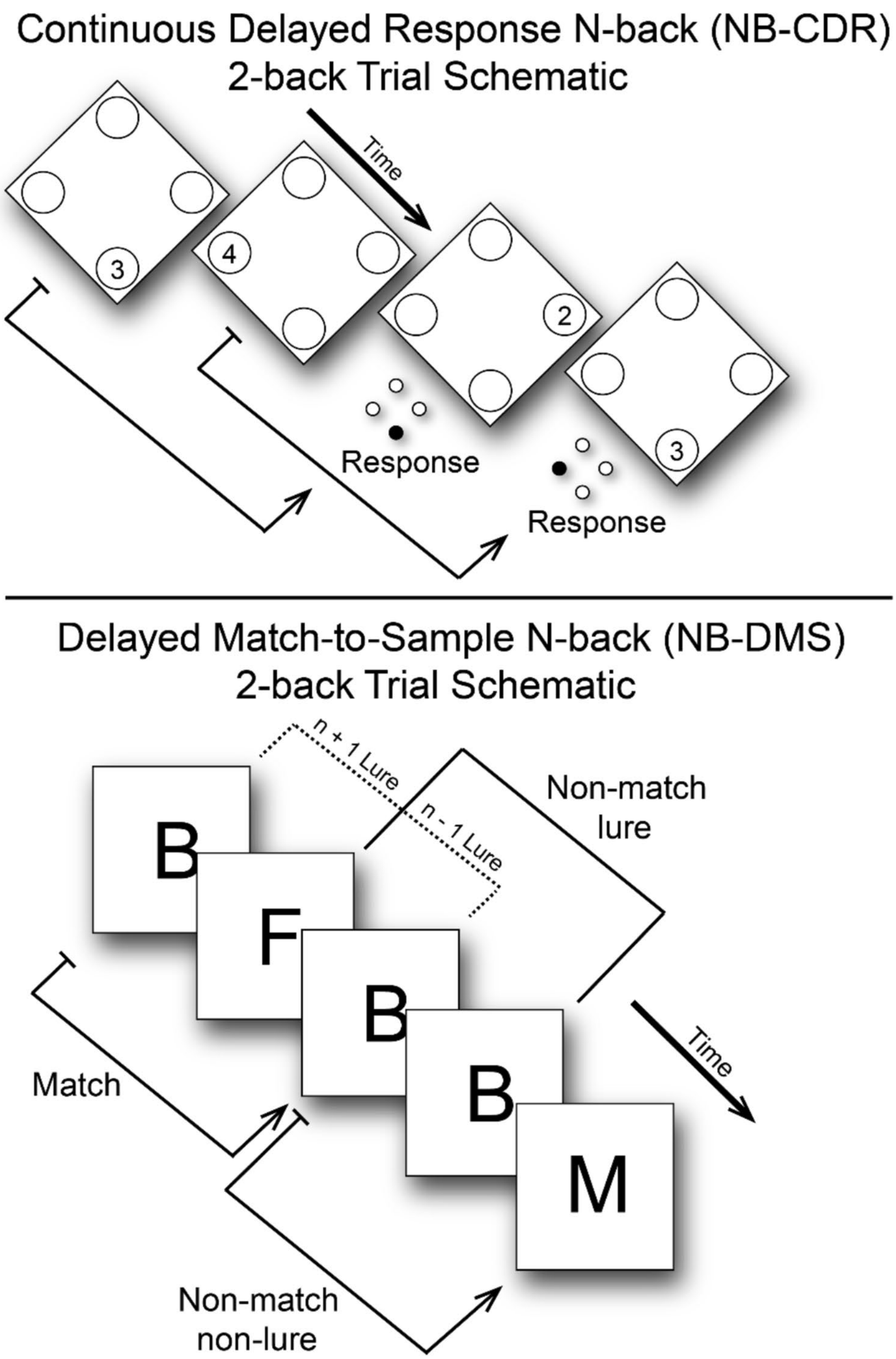


Fig. 2 **Top**: Schematic of an NB-CDR 2-back trial. Stimuli typically consist of both numbers and spatial locations, and on each trial participants must make the response corresponding to the stimulus displayed n trials previously. **Bottom**: Schematic of an NB-DMS 2-back trial, including a recent-probe lure trial. Participants must indicate whether each presented stimulus matches the stimulus presented n trials previously

correlated (correlations ranged from not significant to 0.32), although both types of tasks are theoretically designed to measure the same underlying WM construct (Kane et al., 2007; Redick & Lindsey, 2013; Van Snellenberg et al., 2014a, b). Furthermore, both types of tasks have been found to account for unique variances in fluid intelligence (Kane et al., 2007). This is sometimes interpreted as evidence for the multifaceted nature of WM, such that complex span tasks and n-back tasks require separate components of the WM system (Redick & Lindsey, 2013). While complex span tasks mostly access executive attention processes, n-back tasks are thought to rely on updating processes; that is, at 2-back loads and higher, n-back tasks are relatively unique in requiring both addition and removal of items from the memory set (rather than simple addition in task with serial presentation of targets, such complex span tasks), in addition to maintaining the temporal ordering of stimulus presentation.

Despite the popularity of complex span tasks and n-back tasks, several other tasks are also used in cognitive and psychiatric neuroimaging research to measure WMC. The Sternberg Item Recognition Paradigm (SIRP) is a short-term memory task requiring recognition of (typically) letters or words, but is often used as a WM task, especially when larger target sets (~5 to 8 items) are used (Sternberg, 1966; Rypma et al., 2002). The AX-Continuous Performance Test (AX-CPT) is used to measure context processing and goal maintenance (Redick & Engle, 2011). In this task, participants respond to specific cue-target sequences (e.g., respond to "X" only if it was preceded by "A"), which requires maintenance of the recent trial context in order to respond accurately. Change detection tasks are short-term memory tasks that provide a measure of visual WMC via an estimate of the number of items an individual can hold in short-term memory at a time (Vogel & Machizawa, 2004). Finally, the Self-Ordered WM Task (SOT) requires participants to select all of the items in an array (of ~8 to 12 items) only one time each, in any order, thus requiring participants to remember all previously selected items throughout a trial and providing a fine-grained variation in WM load from 0 up to 8 or 12 items (Van Snellenberg et al., 2014a, b, 2016).

Domain general and domain-specific perspectives of working memory

It has been generally observed that all types of WMC measures are positively intercorrelated (Oberauer et al., 2012). This pattern of positive intercorrelations has led to a domain-general interpretation of WM capacity, suggesting that the psychometric general factor of WM largely reflects properties of a domain-general mechanism, namely executive attention (Engle & Kane, 2004). From this point of view, the variation in WM and its relationship with the broad range of real-world outcomes are primarily driven by individual differences in executive attention. This interpretation is supported by the central role WM plays in higher-order cognitive abilities, such as fluid reasoning (Cowan, 2005). However, correlational evidence has shown that WMC is not always best accounted for by a unitary factor, but instead by multiple correlated lower-level factors that represent domain-specific processes such as verbal/spatial storage, coding, and rehearsal (Oberauer, 2019a, b). These factors indicate that individual differences in WMC may be determined jointly by domain-general cognitive mechanisms, such as executive attention, and domain-specific mechanisms, such as verbal/spatial storage. This view has also been supported by studies testing the cross-domain predictive validity of different complex span tasks, such as verbal and spatial span tasks, in which verbal span tasks were found to predict performance on other verbal measures but not on spatial ability measures (Daneman & Merikle, 1996; Morrell & Park, 1993; Shah & Miyake, 1996).

This theoretical diversity is also reflected in formal attempts to delineate the structure of WM subconstructs. The Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia (CNTRICS) initiative identified seven subconstructs of WM (Barch et al., 2009): Goal Maintenance, Interference Control, Maintenance Over Time, Updating, Strategic Encoding, Long-Term Memory Reactivation, and Capacity (Barch & Smith, 2008). In the NIMH Research Domain Criteria (RDoC) Matrix (2024), four subconstructs of WM were formally encoded: active maintenance, flexible updating, limited capacity, and interference control. In the current literature on individual differences in WM, some research focuses on a single general ability associated with domain-free, attention-related cognitive functions of WM, such as those proposed by Oberauer (2019a) and Engle (2002), while other research attempts to identify more specific cognitive abilities underlying WM capacity, thought to reflect different subconstructs of WM. As a result of these different approaches, several different psychometric models of WM have been proposed. Some models highlight the importance of executive attention as the critical, domain-general factor of individual differences in WM (Engle & Kane, 2004), while other models propose multiple factors, thought to reflect multiple subconstructs of WM (Unsworth et al., 2014). Unfortunately, the debate is still ongoing for the number of factors that best account for individual differences in WM. These different theoretical perspectives underscore the complexity of WM and the need for precise measurement tools and demonstrate the many challenges of studying WM.

Over the years, various WM tasks have been employed across different fields and for different research purposes, leading to inconsistencies in findings. Most WM tasks are inherently complex, and likely measure multiple cognitive processes (Chatham et al., 2011; Unsworth et al., 2009). Furthermore, although different WM tasks, such as n-back and complex span tasks, were conceptually developed to measure the same WM system, they may not reflect the same underlying cognitive processes (Redick & Lindsey, 2013). In addition, the types and formats of the to-be-retrieved stimuli involved in these WM measures are usually different across tasks. Each of these differences in cognitive functions and task properties may have unique associations with cognitive behaviors and deficits, such as those observed in schizophrenia or other psychiatric disorders, thereby influencing research findings based on the selection of test measures and target cognitive behaviors or deficits.

These issues highlight that WM tasks are not functionally equivalent, and that measurements of individual performance across tasks may not reflect the same underlying cognitive processes. This heterogeneity poses significant challenges for researchers attempting to understand the structure of WM in health and disease. As such, WM performance deficits in patients with neurological and psychiatric disease observed using a variety of WM tasks could indeed reflect a diverse variety of interrelated deficits in cognition, each with distinct underlying neurobiological mechanisms. That is, neural correlates of WM deficit, such as those measured using functional neuroimaging or electroencephalography, may depend on the cognitive processes elicited by specific task paradigms, rather than simply reflecting a unitary WM deficit.

In the current study, we sought to address these challenges by investigating a comprehensive set of WM tasks that are commonly used in cognitive psychology and functional neuroimaging research. A battery of WM tasks, as well as other neurocognitive measures, were administered to a large sample of healthy young adults, and a series of factor analyses and structural equation models were conducted to examine the covariance structure of individual differences in WMC. This allowed us to test the unitary general ability view of WM capacity (Engle, 2002) and the nonunitary, multi-faceted view (Kane et al., 2004; National Institute of Mental Health; 2024), and to determine whether the latent structure of WM appears to match the theoretical structures proposed by, e.g., CNTRICS and RDoC. To test these competing accounts, we investigated the latent factor structure of WMC using a variety of cognitive tasks: complex span, NB-DMS, NB-CDR, SIRP, AX-CPT, change detection, SOT, and a subset of neurocognitive tests from the Penn Computerized Neurocognitive Battery (PennCNB) (Gur et al., 2010). The cognitive tasks included from the PennCNB were included in order to capture variance related to other cognitive processes, such as attention, long-term memory recognition, and processing speed. By examining the relationships among these tasks, we aimed to clarify the underlying structure of WM and its components to provide novel insights to the construct validity of WM measures for applied research, such as psychiatry and neuroimaging.

Methods

Participants

A total of 608 participants who were enrolled in undergraduate courses in psychology at Stony Brook University were recruited to participate in this study. Inclusion criteria for participants were 1) ability to use a computer and 2) sufficient English language proficiency to understand task instructions. Participants received course credit for participation. See Table 1 for a demographics summary table. Stony Brook University admissions data indicates that undergraduates have a wide range of cognitive abilities (e.g., the middle 50% of the 2018 freshman class had SAT scores from 1,250 to 1,400, suggesting that 25% of our sample will have scores > 1,400 and 25% will have scores < 1,250), which allowed us to capture meaningful variance in WM subconstructs.

Cognitive task battery

All participants completed a battery of 18 tasks, including 9 WM tasks that are commonly used in psychiatric neuroimaging and cognitive psychology research, 7 tasks from the Penn Computerized Neurocognitive Battery (PennCNB) (Gur et al., 2010), and 2 subtests from the *Wechsler Abbreviated Scale of Intelligence*, 2nd Edition (WASI-II) (Wechsler, 2011). See Table 2 for a summary of the tasks.

For each task, one or more measures were scored. However, for the purposes of this study, only a subset of these measures was selected for further analysis, particularly for exploratory factor analyses (EFA). Specifically, all reaction time (RT) variables, except those from the Motor Praxis Test (MPRAXIS) and Finger-Tapping Task (FTAP), were excluded from the analyses. Preliminary

Table 1. Demographics of the participants recruited at stony brook university for the current study

	N (SD or %)
Age	19.83 (2.39)
Gender	
Male (M)	197 (32.40%)
Female (F)	407 (66.94%)
Not reported (NA)	4 (0.66%)
Race	
White (W)	159 (26.15%)
African American (AA)	62 (10.20%)
Asian (AS)	325 (53.45%)
American Indian (AI)	2 (0.33%)
Pacific Islander (PI)	2 (0.33%)
Multiracial (MR)	29 (4.77%)
Other	24 (3.95%)
Not reported (NA)	5 (0.82%)
Ethnicity	
Hispanic (H)	68 (11.18%)
Non-Hispanic (NH)	458 (75.33%)
Not reported (NA)	82 (13.49%)

Table 2 Description of the cognitive tasks included in the current study

Task Type	Task Name	Description
Common WM Tasks in fMRI Studies	Delayed match-to-sample n-back task (NB-DMS)	Tests working memory through delayed matching of samples with added n-back complexity.
	Continuous delayed response n-back (NB-CDR)	Measures continuous working memory performance with delayed responses under n-back conditions.
	Sternberg item recognition paradigm (SIRP)	Assesses memory for a list of items by having participants indicate whether a probe was in a previously seen list.
	Visual item recognition paradigm (VIRP)	Evaluates visual memory by requiring identification of previously presented visual items.
	AX-Continuous performance test (AX-CPT)	A measure of sustained attention and working memory where participants respond to target sequences presented in a stream of letters.
	Change detection tasks (CD)	Assesses the ability to detect changes to an array of visual stimuli after a brief interruption.
	Self-ordered working memory/pointing task (SOT)	Evaluates working memory through self-generated sequences of responses in a spatial array.
	Operation Span Task (OSPAN)	Combines memory for a sequence of verbal items with a concurrent processing task (solving math operations).
	Symmetry Span Task (SSPAN)	Tests working memory by requiring recall of sequences of spatial locations interspersed with symmetry judgments.
PennCNB Tasks	Motor praxis (MPRAXIS)	Measures motor speed and precision through task-oriented motor movements.
	Penn facial memory delayed (CPFD)	Tests recognition memory for faces, including the ability to remember faces over time (D for delayed).
	Penn number/letter continuous performance test (PCPTNL)	A task to assess sustained and selective attention by monitoring for specific sequences of numbers or letters.
	Penn abstraction, inhibition, and WM (AIM)	Evaluates the ability to abstract concepts, inhibit responses, and utilize working memory.
	Visual object learning delayed (SVOLTD)	Assesses learning and memory for visual objects, including retention over a delay (D for delayed).
	Digit symbol substitution task (DIGSYM)	Measures speed and accuracy of digit-symbol coding under timed conditions.
	Finger-tapping task (FTAP)	Assesses motor speed through repetitive finger tapping.
WASI-II Tasks	Matrix Reasoning (WASI-Matrix)	Tests non-verbal abstract problem solving, logical reasoning, and pattern recognition.
	Vocabulary (WASI-Verbal)	Measures word knowledge and verbal concept formation.

factor analyses indicated that RT measures across all tasks consistently loaded onto a single RT factor, distinct from other task variables.

That said, RT variables from MPRAXIS and FTAP were retained but only as part of a residualization procedure (see details below). For all other tasks, a single non-RT performance measure was selected based on two criteria: (1) its common usage in the literature, and/or (2) its observed split-half reliability. Thus, for each task, a single accuracy-based measure with acceptable reliability was retained for analysis. Details on the split-half reliability analysis are provided in the Supplementary Information ("Reliability Analysis").

The following section describes the specific tasks administered and their corresponding primary measures used in this study's analyses.

Working memory tasks

Operation span (OSpan) The Operation Span (OSpan) Task consisted of 15 items. Each item had a varying number of alternative processing and storage components, where the processing components were always followed by a storage component. In the OSpan Task, the processing component consisted of a mathematical equation (e.g., "$(2 \times 2) + 1 = 5$") that was followed by the storage component, a letter. For each set of processing and storage stimuli, subjects had to judge whether the mathematical equation was correct and then memorize the letter. In total, there were 15 items presented across 3 blocks, and the item sizes varied from 3 to 7 sets of stimuli. Each item size was randomly presented once in each of the 3 blocks. The primary outcome measure was the Partial Unit Score (OSpan PU), which reflects the proportion of correctly recalled letters across trials.

Symmetry span (SSpan) The Symmetry Span (SSpan) Task consisted of 12 items. Each item had a varying number of alternative processing and storage components, where the processing components were always followed by a storage component. In the Symmetry Span Task, the processing component consisted of a figure that was followed by the storage component, a random colored cell in a 4 × 4 grid. For each set of processing and storage stimuli, subjects had to judge whether the figure was symmetrical and then memorize the position of the colored cell in the 4 × 4 grid. In total, there were 12 items presented across 3 blocks, and the item sizes varied from 2 to 5 sets of stimuli. Each item size was randomly presented once in each of the 3 blocks. The primary outcome measure was the Partial Unit Score (SSpan PU), representing the proportion of correctly recalled locations in sequence.

AX continuous performance test (AXCPT) The AX Continuous Performance Test (AXCPT) task consisted of 20 trials in each run. Each trial consisted of two stimuli: a cue and a probe. Cues were either the letter A (A-cue) or any other letter except X (collectively, B-cues), while probes were either the letter X (X-probe) or any other letter except A (collectively, Y-probes). For each trial, subjects pressed a button to indicate if the cue-probe combination was the target (AX) or a non-target (any other combination). Stimuli were presented for 500 ms, and the delay between cue and probe was randomly jittered from 2–4 s (mean of 3 s). The ITI was randomly jittered from 4–6 s (mean of 5 s). Each run of the task consisted of 70% AX trials, 10% BX trials, 10% AY trials, and 10% BY trials. The primary outcome measure was AXCPT d′-Context (Gonthier et al., 2016), a signal detection measure quantifying the ability to discriminate between AX and BX trials.

Change detection Trials consisted of a target and probe presentation. Targets were two, four, six, or eight colored squares (red, green, blue, yellow, white, or black; no color appeared more than twice in any target), presented in one of four quadrants around a fixation crosshair. Stimuli were balanced across all quadrants, ensuring even distribution. The probe appeared in the same location as a previous stimulus, and participants indicated whether the probe's color matched the original square's color. Targets were presented for 500 ms, followed by a 1 s delay. Probes were presented for 5 s, during which participants made a target/non-target response. The ITI was 1 s. The primary outcome measure was d′.

N-Back (Continuous Delayed Response; NBCDR) The Continuous Delayed Response N-Back Task (NBCDR) involved presenting single stimuli (numbers 1–4) at fixed locations within a diamond arrangement. Each stimulus was displayed for 500 ms, followed by a 1,500 ms delay. A number at the top of the screen indicated the condition (1-back or 2-back). Participants responded by pressing a corresponding button on a response pad when the stimulus location matched that of the item presented one or two trials earlier. The primary outcome measure was NBCDR Accuracy, the proportion of correct responses across all conditions.

N-Back (Delayed Matching-to-Sample; NBDMS) In the Delayed Match-to-Sample N-Back Task (NBDMS), single non-vowel letters were presented for 2 s each, followed by a 2 s interstimulus interval (ISI), in blocks of 10 trials. Each block was followed by a 7 s delay before a cue indicated whether it was a 1-back or 2-back block. During these conditions, participants pressed a button when the letter matched the one presented one or two trials previously. Each block contained 50% match trials and 50% non-match trials. Half of the nonmatch trials were lure trials, which contained the target but were not accurately 1 or 2 trials back for the 1- or 2-back conditions, respectively. The primary outcome measure was NBDMS Accuracy, the proportion of correctly identified matches across all conditions.

Sternberg item recognition paradigm (SIRP) The Sternberg Item Recognition Paradigm (SIRP) consisted of trials that included a target and probe presentation. Target sets contained two, four, six, or eight uppercase consonants, displayed for 4 s, followed by a 3–7 s jittered delay (mean of 5 s). A probe letter was then presented for 4 s, and participants indicated whether the probe matched any of the target items. Half of the nonmatch trials (25% of total trials) were recent probes, meaning the probe was from the previous trial but not the current trial. Each run consisted of 16 trials (4 of each load in pseudorandomized order). The primary outcome measure was d′.

Visual item recognition paradigm (VIRP) The Visual Item Recognition Paradigm (VIRP) followed the same design as the SIRP except that 1) only two, four, or six stimuli are used in each target array; 2) the stimuli are three-dimensional drawings of objects (identical to the SOT task, described below); and 3) there are 21 trials per run (seven of each load, in pseudorandom order). The recent probe manipulation was identical to the SIRP. The primary outcome measure was d′.

Self-ordered working memory task (SOT) The Self-Ordered Pointing Task (SOT) involved 24 trials, each lasting 72 s and consisting of eight steps (9 s each), with an 8 s delay between trials. Participants viewed a 3 × 3 grid of eight three-dimensional objects and a fixation cross and were required to select each stimulus once and only once per trial. Stimuli were randomly rearranged after each selection. In each trial, this continued until all eight objects were chosen. The primary outcome measure was SOT Capacity Score (*k*), which estimates participants' working memory capacity (Van Snellenberg et al., 2014a, b).

Neurocognitive tasks from the PennCNB battery

Penn continuous performance test - Number letter (PCPT-NL) The Penn Number/Letter Continuous Performance Test (PCPT-NL) is a measure of visual attention and vigilance. Participants viewed a sequence of red vertical and horizontal lines that flashed within a digital numeric frame (resembling a digital clock). Their task was to press the spacebar whenever the displayed lines formed complete numbers or complete letters. The task consisted of two conditions: one where participants identified numbers, and one where they identified letters, each lasting 3 min. Each stimulus flashes for 300 ms and a blank page is then displayed for 700 ms, giving the participant 1 s to respond to every trial. The primary outcome measure was d'.

Abstraction, inhibition, and memory (AIM) The Penn Abstraction, Inhibition, and Memory (AIM) Task assessed concept formation and working memory through a categorization task. Participants were shown a single object and asked to determine which of two object pairs it best matched, based on color or shape. In the working memory condition, the object disappeared briefly before the choice was presented, requiring participants to maintain it in memory. The primary outcome measure was the AIM Sum Score for Working Memory Trials, reflecting overall task performance across these trials.

Penn facial memory test - delayed (CPFD) The Penn Facial Memory Test (CPFD) assessed delayed recognition memory for faces. In the learning phase, participants were shown 20 unfamiliar faces to remember. In the delayed recall phase, they were presented with 40 faces, consisting of the original 20 and 20 novel distractor faces. Participants rated their confidence in recognizing each face using a 4-point scale ("definitely yes"to"definitely no"). The primary outcome measure was d'.

Short visual object learning test - delayed (SVOLT-D) The Short Visual Object Learning Test (SVOLT-D) was designed as a spatial analog to the California Verbal Learning Test, assessing visual learning and memory. Participants initially viewed 10 three-dimensional Euclidean shapes in a study phase. Later during the delayed recall, they were presented with 20 shapes—a combination of previously studied shapes and novel distractors—and indicated whether each had been seen before. The primary outcome measure was d'.

Digit symbol substitution task (DIGSYM) The Digit Symbol Substitution Task (DIGSYM) assessed processing speed and cognitive efficiency (Knowles et al., 2012). Participants referred to a legend linking digits (1–9) to specific symbols and determined whether a displayed digit-symbol pair was correct. A total of 54 trials were presented randomly, and participants had 1.5 min to complete as many trials as possible. The primary outcome measure was the Digit Symbol Score, reflecting the number of correct responses within the time limit.

Motor praxis test (MPRAXIS) The Motor Praxis Test (MPRAXIS) measured sensory-motor coordination and computer mouse proficiency. Participants were required to move the cursor over a shrinking green box and click on it as it appeared in different screen locations. This task was included to measure motor skill and proficiency with a computer mouse before proceeding to other computerized tasks. The primary outcome measure was MPRAXIS Trial 2 Median Reaction Time, quantifying the median time taken to respond to the shrinking target.

Finger-tapping task (FTAP) The Finger-Tapping Task (FTAP) assessed manual dexterity and motor speed. Participants tapped the space bar as many times as possible in 10-s trials, alternating between the dominant and nondominant hands. The task included five trials per hand, with two additional 5-s practice trials before the test. The primary outcome measure was the FTAP Total Score, which summed the total number of taps across both hands.

General cognitive ability

Wechsler abbreviated scale of intelligence - second edition (WASI-II) The WASI-II was administered as an estimate of general cognitive ability. Two subtests were used: Vocabulary T-Score (WASI Verbal) and Matrix Reasoning T-Score (WASI Matrix).

Study procedures

All procedures were approved by the Stony Brook University Institutional Review Board, and all individuals provided written informed consent prior to their participation in the study. Participants completed the cognitive task battery in two sessions lasting less than 2 h each. The placement of tasks in each session, as well as the order of tasks within sessions, was pseudorandomized across participants. Tasks within the PennCNB were presented in a fixed order (in the order shown in Table 2), and two subscales of the WASI-II were conducted during a break in the PennCNB tasks that occurred between the AIM and SVOLT-D (Table 2). All other tasks and the block of tasks comprising the PennCNB and WASI-II were ordered according to a Latin squares design, except in cases where the Latin square could not produce two sessions of less than 2 h (i.e., some task orderings would produce one session of greater than two hours, and so were excluded). Participants were offered an optional 5-min break every 30 min or less, which was included in the 2-h session time limit. Participants were also given the opportunity to practice each of the WM tasks until they indicated that they fully understood the task instructions immediately prior to beginning each task. All of the WM tasks were presented using Presentation software (Neurobehavioral Systems, Inc., Berkeley, CA), with the exception of the two complex span tasks, which were presented in E-Prime 3.0 (Psychology Software Tools, Inc., Pittsburgh, PA) (Oswald et al., 2015).

Data preparation

Outlier analysis

We adopted a systematic, a-priori approach to handle outliers for the cognitive measures. Specifically, we identified univariate outliers as those with an absolute z-score greater than 3.5, indicating a distance of more than 3.5 standard deviations from the mean. A total of 65 univariate outliers were flagged; however, these univariate outliers were not entirely removed from the analysis. Instead, only their specific task scores that deviated from the mean more than 3.5 standard deviations were regarded as missing and excluded in later analyses, allowing us to retain as much usable data as possible while minimizing potential bias from extreme values. Re-running the analyses with the outliers included revealed no meaningful changes to the findings, suggesting that this approach did not materially alter our results. Raw results with and without the outliers are available at https://osf.io/cvfta/. No multivariate outliers were excluded from the analysis, instead, robust estimations were applied to account for the violation of multivariate normality (see "Data analysis"). During data residualization (see "Residualization"), participants who failed to complete the FTAP and MPRAXIS tests in Penn tasks were excluded from further analyses because individual scores on these two tasks were used to residualize subjects' performance on all other selected tasks. As a result, 38 subjects were excluded because of their missing FTAP and MPRAXIS scores, reducing the sample size from 608 to 570.

Residualization

Before the EFA, we conducted a data residualization process to control for motor skill influence associated with different input devices across different tasks (keyboard vs. mouse). In general, this controls for the effects of covariate variables when examining the relationship between other target variables of interest. In tasks where participants respond via distinct motor modalities, differences in motor skills can inadvertently influence performance on cognitive tasks. Therefore, by using performance on the Finger Tapping (FTAP; keyboard-based) and Mouse Praxis (MPRAXIS; mouse-based) tasks as covariates, we attempted to statistically remove ("residualize") any variance in these cognitive measures due to motor demands prior to examining the intercorrelations among WM tasks.

Specifically, each cognitive measure that was administered under mouse input (all 5 Penn Tasks other than MPRAXIS and FTAP themselves) were regressed (residualized) on MPRAXIS to obtain scores that were free from mouse-based motor skill influence. Conversely, all nine keyboard-based WM measures were regressed (residualized) on FTAP to obtain scores that were free from keyboard-based motor influence. The two WASI tests were not included for the residualization process as these were traditional paper-and-pencil measures. In total, 14 measures were residualized. By using these residualized scores in the later factor analyses, we could more accurately identify underlying WM constructs rather than capturing "noise" variance from motor abilities from different input types.

Data analysis

The 14 residualized variables for 570 subjects were included in the following analyses. A series of EFAs was conducted to investigate the factor structure of WM. Full information maximum likelihood estimation (FIML) was used to calculate the covariance matrix for the EFAs to account for missing values. The sample of responses on the 14 measures did not follow a multivariate normal distribution, according to the Henzer-Zirkler's test (Henze & Zirkler, 1990): HZ = 1.08, p <.001. Thus, weighted least square (WLS) estimation in all EFA analyses so that the non-normal distributions of the items were accounted for, and the oblique ("oblimin") rotation was used so that the latent factors from the EFAs were all correlated.

Prior to the EFAs, the assumption of sampling adequacy was checked using the Kaiser-Meyer-Olkin (KMO) sampling adequacy test. Both overall and variable-level MSAs were estimated. Based on the KMO results, the current sample of 14 measures was an adequate sample of manifest variables for latent variable analyses. The overall MSA (measures of sampling adequacy) was 0.87, and variable-level MSAs ranged from 0.79 to 0.93. No arbitrary cut-off was determined but MSAs larger than 0.7 are considered more than "middling" (Kaiser & Rice, 1974). A parallel analysis (Horn, 1965) was conducted to estimate the number of potential factors to be extracted. The eigenvalues from observed data were compared to the average of 1,000 eigenvalues from random simulations of datasets with the same size and structure as the observed data.

For all the EFA solutions, a variety of indicators were evaluated to determine the proper factor solution, including the factor structures, fit indices such as Tucker Lewis Index of factoring reliability (TLI) and root mean square error of approximation (RMSEA), cumulative variances explained by the solution, factor loadings, and extracted communalities for observed variables (h^2). Cumulative variances are the estimated proportion of variances that the EFA solution can explain in the data. Factor loadings represent how well the items reflect the latent construct, with a standardized loading >.30 being acceptable. Extracted communalities for items represent how much of the variance in the item is attributed to the target factor solution, with higher communalities being preferred.

After an EFA solution that identifies the subconstructs underlying WM performance was determined, a series of regression analyses were conducted in structural equation models (SEMs). In these SEM models, the latent factors representing the identified subconstructs were used as predictors to subjects' intellectual testing performance, measured by the WASI-II verbal and matrix reasoning subtests. Each of the latent subconstructs was assumed to uniquely explain the variances in subjects' performance on WASI-II verbal and matrix reasoning subtests. The regression coefficients, as well as the overall goodness of fit of the SEM models, were evaluated. The fit of each model is evaluated with the following set of test statistics and fit indices: χ^2, Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), and Akaike Information Criterion (AIC). The criteria for an "acceptable" model fit were based on the following cut-off values, recommended by Schreiber et al. (2006) and Kline (2016): $p > .05$, $\chi^2/df \leq 3$, CFI $\geq .95$, RMSEA $\leq .06$, SRMR $\leq .08$. For AIC there is no cutoff value; lower values indicate better fit. Robust maximum likelihood estimation was used for all SEM models, such that chi-square statistics and all corresponding fit indices were corrected using the Huber-White correction (Rosseel, 2012). Similar to the EFAs, all missing values were handled via FIML estimation.

Results

Summary statistics for the primary 18 measures (not residualized) are reported in Table 3. The correlation matrix of the residualized measures is reported in Table 4.

Exploratory factor analysis

The result of the parallel analysis indicated that a three-factor solution was optimal (Fig. 3). Hence, a 3-factor, oblique, weighted least square EFA was performed. The findings from this initial EFA suggested that a three-factor model explains 36% of the total variance across all 14 residualized measures, with a Tucker-Lewis Index (TLI) of.87 and a Root Mean Square Error of Approximation (RMSEA) of.063, 90% CI [.052,.073].

The factor structure revealed by this analysis (EFA1) is presented in Table 5, demonstrating that most of the WM measures loaded highly on the first factor. The two n-back tasks were predominantly loaded onto the second factor, while the AIM measure was uniquely loaded onto the third factor. Among the 14 measures, OSpan did not load well on any of the three factors, with standardized factor loadings ranging from 0.03 to 0.25. Standardized factor loadings and the correlations among factors are presented in Table 5. All factor loadings <.3 were omitted from the table.

Given concerns about the third factor, where only the AIM was loaded significantly with a loading close to 1, and considering the low loading of the OSpan measure, a 4-factor EFA solution was also conducted (EFA2). Results of this solution indicated that the 4-factor solution accounted for 40% of the total variance, along with improved fit indices: TLI =.95 and RMSEA =.04, 90% CI [.026,.053]. Similar to EFA1, the majority of the measures were loaded onto the first factor. However, the two complex span measures were distinctively loaded onto a separate factor (F2), and the n-back measures onto another (F3), illustrating a clearer distinction among the task types. The AIM measure was still loaded onto its own factor.

Based on the results from EFA1 and EFA2, the AIM was excluded from the selected measures, and a 3-factor oblique solution was retained. This was confirmed by another parallel analysis of the 13 remaining measures. In this solution (EFA3), the 3-factor structure accounted for 36% of the total variance, with TLI =.96 and RMSEA =.038, 90% CI [.024,.052]. The factor structure was similar to EFA2 (Table 6), except the fourth factor and AIM were no longer included. The three factors were moderately correlated, with correlations ranging from 0.4 to 0.53.

Given the 3-factor structure, we interpret the first factor with the majority of measures except n-back and Span tasks as a "Storage" factor, the second factor with the two span tasks as an "Executive Attention" factor, and the third factor with the two n-back tasks as an "Updating" factor.

Table 3 Descriptive statistics and split-half reliability estimates for the selected measures

Measures	N	Mean	SD	Skew	Kurtosis	Reliability
OSpan	574	0.76	0.21	−0.99	0.41	0.66
SSpan	571	0.64	0.23	−0.59	−0.34	0.62
AXCPT	559	2.38	1.17	−1.34	1.44	0.90
CD	561	1.72	0.78	−0.36	0.27	0.89
NB-CDR	568	0.66	0.28	−0.82	−0.44	0.97
NB-DMS	524	0.84	0.18	−1.73	2.16	0.98
SIRP	558	2.76	0.89	−0.93	1.28	0.87
VIRP	565	1.48	0.72	−0.25	0.15	0.68
SOT	558	5.87	1.38	−1.78	3.05	0.86
PCPTNL	567	3.65	0.77	−0.15	−0.15	0.90
AIM	569	25.10	2.91	−0.78	0.42	0.65
CPFD	552	2.24	0.76	−0.12	−0.17	0.68
DIGSYM	571	55.48	10.44	0.04	−0.11	NA
SVOLTD	568	1.64	0.86	−0.10	−0.53	0.57
MPRAXIS	573	574.21	89.95	1.55	5.15	NA
FTAP	570	112.75	37.34	8.30	94.20	NA
WASI Verbal	577	55.63	10.31	−0.35	0.80	NA
WASI Matrix	577	50.00	9.23	−0.08	0.88	NA

Note. Ospan = Operation Span Partial Unit Score; Sspan = Symmetry Span Partial Unit Score; AXCPT = AX Continuous Performance Test Context d'; CD = Change Detection d'; NBCDR = N-Back (Continuous delayed response) Accuracy; NBDMS Accuracy = N-Back (Delayed Matching-to-Sample) Accuracy; SIRP = Sternberg Item Recognition Paradigm d'; VIRP = Visual Item Recognition Paradigm d'; SOT = Self-Ordered Working Memory/pointing task Capacity Score; PCPTNL = Penn Continuous Performance Test Number Letter d'; AIM = Abstraction, Inhibition, and Memory Sum Score of Working Memory Trials; CPFD = Penn Facial Memory Test (Delayed) d'; DIGSYM = Digit Symbol Score; SVOLTD = Short Visual Object Learning Test (Delayed) d'; MPRAXIS = Motor Praxis Test Trial 2 Median Reaction Time; FTAP = Finger-tapping task total score; WASI Verbal = Wechsler Abbreviated Scale of Intelligence (2nd Edition) Vocabulary T Score; WASI Matrix = Wechsler Abbreviated Scale of Intelligence (2nd Edition) Matrix Reasoning T Score. NA: Not Applicable. Split-half Reliability was not a properly defined metric for DIGSYM, MPRAXIS, and FTAP. For the two WASI measures, only subtest scores were available and used in the current study

Regression analyses using structural equation models

Based on the EFA results, a series of Structural Equation Models (SEMs) were specified to examine the relationship between the WM factors and intelligence. More specifically, the three factors identified in the EFAs were used to predict scores on a two-subtest version of the WASI-II (Wechsler, 2011), which consisted of a verbal subtest (WASI-Verbal) and a matrix reasoning subtest (WASI-Matrix). In each of the following SEMs, the 3 WM components (Storage, Executive Attention, Updating) were specified as predictor variables, and the two intelligence measures (WASI-Verbal, WASI-Matrix) were specified as outcome variables.

According to the EFA result, the Storage factor consists of most of the selected measures and, conceptually, may stand as the factor that taps more cognitive processes and therefore better accounts for individual differences in WM compared with the other two factors. Thus, for the first SEM, all 3 WM components were included, but only Storage was specified as a predictor of intelligence (Fig. 4). Results indicated an acceptable model fit: $\chi^2(86) = 197.65$, $p < .001$; CFI =.92, TLI =.91; RMSEA =.05, 90% CI [.04,.06]; SRMR = 0.04; AIC = 22318.66, BIC = 22531.6. According to the model, the Storage component of WM is a statistically significant predictor of both WASI-Matrix ($\beta_m = .38$, $p < .001$) and WASI-Verbal ($\beta_v = .21$, $p < .001$). Overall, the Storage component explained approximately 14% of the variance in WASI-Matrix and approximately 4% of the variance in WASI-Verbal.

Although in the current factor model, the storage factor consists of the majority of the selected measures, the factor is only moderately correlated with the other two factors that consist of the WM tasks that are more commonly used in cognitive psychology/neuroscience literature. Thus, in the second SEM, all three WM factors (Storage, Executive Attention, and Updating) were included as predictors of intelligence (Fig. 5). Results indicated an acceptable model fit: $\chi^2(82) = 150.04$, $p < .001$; CFI =.95, TLI =.94; RMSEA =.04, 90% CI [.03,.05]; SRMR = 0.04; AIC = 22278.77, BIC = 22509.09. In this model, when the Executive Attention and Updating factors were included, Storage was no longer a significant predictor of WASI Matrix ($\beta_m = -.06$, $p = .486$) or WASI Verbal ($\beta_v = .16$, $p = .088$). Instead, both

Table 4 Correlation matrix for the selected residualized measures

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
OSpan	-														
SSpan	0.41	-													
AXCPT	**0.06**	**0.10**	-												
CD	0.21	0.38	0.29	-											
NB-CDR	0.19	0.30	0.25	0.32	-										
NB-DMS	0.13	0.22	0.19	0.32	0.47	-									
SIRP	0.23	0.23	0.27	0.40	0.25	0.30	-								
VIRP	0.19	0.28	0.29	0.37	0.25	0.24	0.42	-							
SOT	0.20	0.32	0.29	0.34	0.27	0.34	0.31	0.41	-						
PCPTNL	0.25	0.24	0.26	0.35	0.26	0.27	0.30	0.33	0.31	-					
AIM	0.14	0.19	0.12	0.29	0.21	0.18	0.11	0.24	0.20	0.21	-				
CPFD	**0.04**	0.10	0.21	0.15	**0.08**	**0.09**	0.20	0.23	0.27	0.19	0.15	-			
DIGSYM	0.23	0.30	0.16	0.32	0.20	0.20	0.28	0.30	0.32	0.30	0.23	0.19	-		
SVOLTD	0.13	0.28	0.17	0.26	0.17	0.17	0.24	0.24	0.30	0.23	0.16	0.26	0.30	-	
WASI Verbal	**−0.03**	**0.04**	0.14	0.12	0.16	0.23	0.16	0.09	0.14	**0.05**	0.18	0.15	0.09	0.10	-
WASI Matrix	0.21	0.29	0.15	0.28	0.31	0.27	0.16	0.20	0.23	0.16	0.19	**0.05**	0.10	0.18	0.11

Note. All missing values were pairwisely addressed in the correlation matrix. Non-significant ($p > .05$) correlations were in **bold.** Ospan = Operation Span Partial Unit Score; Sspan = Symmetry Span Partial Unit Score; AXCPT = AX Continuous Performance Test Context d'; CD = Change Detection d'; NBCDR = N-Back (Continuous delayed response) Accuracy; NBDMS Accuracy = N-Back (Delayed Matching-to-Sample) Accuracy; SIRP = Sternberg Item Recognition Paradigm d'; VIRP = Visual Item Recognition Paradigm d'; SOT = Self-Ordered Working Memory/pointing task Capacity Score; PCPTNL = Penn Continuous Performance Test Number Letter d'; AIM = Abstraction, Inhibition, and Memory Sum Score of Working Memory Trials; CPFD = Penn Facial Memory Test (Delayed) d'; DIGSYM = Digit Symbol Score; SVOLTD = Short Visual Object Learning Test (Delayed) d'; WASI Verbal = Wechsler Abbreviated Scale of Intelligence (2nd Edition) Vocabulary T Score; WASI Matrix = Wechsler Abbreviated Scale of Intelligence (2nd Edition) Matrix Reasoning T Score

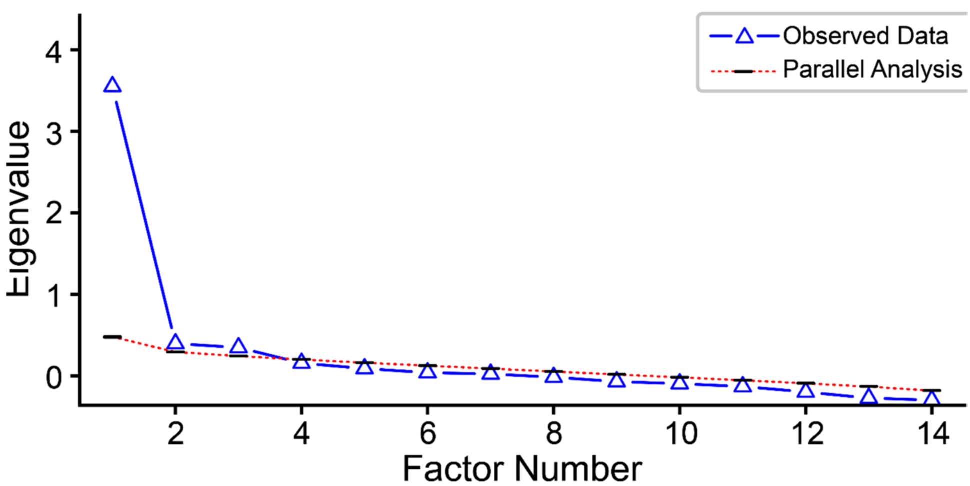


Fig. 3 Parallel analysis with scree plot for the 14 residualized measures. Eigenvalues from the observed data were presented in triangles and solid lines; average eigenvalues from the 1000 simulations were presented in dashed lines. Error bars represent the standard errors of the eigenvalues from the simulations

Executive Attention and Updating were significant predictors of both WASI-Matrix and WASI-Verbal. Overall, the model explained approximately 22% of the variance in the WASI-Matrix subtest and approximately 10% in WASI Verbal.

The results from SEM2 suggest that the relationship between the Storage component of WM and intelligence may be mediated by Executive Attention and Updating. To test this hypothesis, a third SEM was specified to test for mediation. In this SEM, Executive Attention and Updating

Table 5 The Oblique 3-Factor Exploratory Factor Model (EFA1) and 4-Factor Model (EFA2) on the 14 Residualized WM Measures

	EFA1			EFA2			
Task	F1	F2	F3	F1	F2	F3	F4
PCPTNL	0.45			0.40			
AIM			0.99				0.96
CPFD	0.52			0.52			
DIGSYM	0.50			0.40			
SVOLTD	0.51			0.43			
OSpan					0.47		
SSpan	0.35				0.79		
AXCPT	0.40			0.48			
CD	0.46			0.38			
NBCDR		0.66				0.65	
NBDMS		0.57				0.66	
SIRP	0.55			0.53			
VIRP	0.63			0.61			
SOT	0.59			0.53			
Proportion Variance	0.19	0.07	0.07	0.17	0.08	0.08	0.07
Cumulative Variance	0.19	0.26	0.33	0.17	0.25	0.33	0.40
Factor Correlations							
F2	0.55	-		0.47	-		
F3	0.34	0.24	-	0.53	0.40	-	
F4	-	-	-	0.33	0.25	0.25	-

PCPTNL = Penn Continuous Performance Test Number Letter d'; AIM = Abstraction, Inhibition, and Memory Sum Score of Working Memory Trials; CPFD = Penn Facial Memory Test (Delayed) d'; DIGSYM = Digit Symbol Score; SVOLTD = Short Visual Object Learning Test (Delayed) d'; Ospan = Operation Span Partial Unit Score; Sspan = Symmetry Span Partial Unit Score; AXCPT = AX Continuous Performance Test Context d'; CD = Change Detection d'; NBCDR = N-Back (Continuous delayed response) Accuracy; NBDMS Accuracy = N-Back (Delayed Matching-to-Sample) Accuracy; SIRP = Sternberg Item Recognition Paradigm d'; VIRP = Visual Item Recognition Paradigm d'; SOT = Self-Ordered Working Memory/pointing task Capacity Score

were both specified as mediators of the relationship between Storage and both WASI-Matrix and WASI-Verbal (Fig. 6). Results indicated an acceptable model fit: $\chi^2(83) = 151.41$, $p <.001$; CFI =.95, TLI =.94; RMSEA =.04, 90% CI [.03,.05]; SRMR = 0.04; AIC = 22,278.19, BIC = 22,504.17. The results support the hypothesis that Executive Attention and Updating fully mediate t6 he relationship between Storage and the two WASI measures. Indeed, the regression coefficients of Storage on WASI-Matrix and Verbal were both not significant: $\beta_m = -.09$, $p =.349$; $\beta_v =.16$, $p =.109$.

Posteriori confirmatory modeling with primary and secondary memory as separate factors

In some previous research, such as Unsworth et al. (2014) and Robison et al. (2024), primary and secondary memory components were distinguished as reflections of separate aspects of working memory. Therefore, we conducted additional analyses to explore a potential 4-factor solution in which primary and secondary memory are distinct latent variables. In more complex EFA solutions (e.g., a 4-factor and a 5-factor solution), each additional factor was effectively driven by a single task (SVOLT-D for the 4-factor solution and CPF-D for the 5-factor solution), meaning that the extra factors were capturing primarily task-specific variance.

Nevertheless, we tested a confirmatory factor model specifying two distinct storage components, labeled "Primary" and "Secondary," with SVOLT-D and CPF-D forming the Secondary (Memory) factor. This 4-factor measurement model demonstrated an improved fit compared with an original 3-factor CFA model, $\Delta\chi^2(3) = 13.83$, $p <.01$, with minor improvements observed across all fit indices (CFI from.95 to.96, TLI from.94 to.95, RMSEA from.045 to.041, and SRMR from.037 to.034).

To further investigate this structure, we conducted an additional SEM analysis (SEM4), specifying Primary memory and Secondary memory as separate but correlated factors, mirroring the approach used in SEM2, where all WM factors predicted WASI-Matrix and WASI-Verbal. The results did not reveal any substantive differences (Fig. 7). Neither "Primary" factor nor "Secondary" factor was a

Table 6 The Oblique 3-Factor Exploratory Factor Model (EFA3) on the 13 Residualized WM Measures

	F1	F2	F3	h^2
PCPTNL	0.42			0.29
CPFD	0.55			0.21
DIGSYM	0.42			0.28
SVOLTD	0.44			0.23
OSpan		0.45		0.24
SSpan		0.82		0.68
AXCPT	0.47			0.25
CD	0.39			0.39
NBCDR			0.65	0.46
NBDMS			0.65	0.45
SIRP	0.50			0.34
VIRP	0.61			0.40
SOT	0.54			0.40
Proportion Var	0.18	0.09	0.09	
Cumulative Var	0.18	0.27	0.36	
Factor Correlations				
F2	0.46	-		
F3	0.53	0.40	-	

PCPTNL = Penn Continuous Performance Test Number Letter d'; CPFD = Penn Facial Memory Test (Delayed) d'; DIGSYM = Digit Symbol Score; SVOLTD = Short Visual Object Learning Test (Delayed) d'; Ospan = Operation Span Partial Unit Score; Sspan = Symmetry Span Partial Unit Score; AXCPT = AX Continuous Performance Test Context d'; CD = Change Detection d'; NBCDR = N-Back (Continuous delayed response) Accuracy; NBDMS Accuracy = N-Back (Delayed Matching-to-Sample) Accuracy; SIRP = Sternberg Item Recognition Paradigm d'; VIRP = Visual Item Recognition Paradigm d'; SOT = Self-Ordered Working Memory/pointing task Capacity Score

significant predictor of the two WASI measures (Fig. 7). These findings suggest that while distinguishing between two storage components may slightly improve model fit at the measurement level of a confirmatory model, this differentiation does not appear to meaningfully impact their predictive relationships with broader cognitive abilities in the current dataset.

Discussion

This study aimed to investigate the latent structure of WM using a large sample and a comprehensive set of tasks. The exploratory factor analysis (EFA) identified a three-factor model consisting of Storage, Executive Attention, and Updating components. Subsequent structural equation modeling (SEM) revealed that, although Storage predicts the performance of the two measures of intelligence by itself, the association was no longer significant after Executive Attention and Updating were included in the prediction. Executive Attention (complex span tasks) and Updating (n-back tasks, NB-CDR and NB-DMS) were moderately correlated with each other and with the Storage factor, and both factors explained a unique amount of variance in the intelligence measures. The three-factor model supports the notion that WM is not a unitary construct but rather comprises multiple components that contribute uniquely to cognitive performance.

The Storage factor appears to capture a considerable number of both elementary and higher-order cognitive processes that contribute to WM task performance, including measures of sustained attention (PCPT-NL, a continuous performance test), long-term visual recognition (CPF-D and SVOLT-D), processing speed (DIGSYM), context maintenance (AX-CPT), and short-term or WM capacity (change detection, SIRP, VIRP), as well as a complex WM task that likely taps a number of these processes (SOT). In contrast, Executive Attention appears to require the ability to maintain the identity and sequential ordering of target items in the face of ongoing intra-domain distraction from a concurrent processing task. Updating, measured by n-back tasks, captures the ability to revise and refresh information in WM (these are the only tasks in our battery that require concurrent addition *and* removal of items from the memory set, without a full "reset" of target material), and potentially also the ability to temporally tag the contents of WM (so that the correct order of stimuli can be maintained) or to resolve interference from recent-negative lure trials (although lure trials were also included in the SIRP and VIRP tasks).

These results demonstrate that both complex span tasks and n-back tasks have unique properties distinct from each other and from other WM tasks. Although all three factors were relatively highly intercorrelated ($r = 0.48–0.65$), indicating that there is substantial shared variance across all three tasks, when the three factors were considered in a single model, only the Updating and Executive Attention factors had unique shared variance with our measures of intelligence. This suggests that both complex span and n-back tasks may have a closer relationship to real-world cognitive performance than other WM tasks; while the Storage factor is still related to these measures, both n-back and complex span tasks can explain that same variance in intelligence, while also each contributing unique additional explanatory power. This evidence arguably helps cement complex span and n-back tasks as the best options for behavioral or neuroimaging studies of WM, respectively, at least when a closer relationship to measures of intelligence is desired (as it arguably would be for most studies of neuropsychiatric patient groups, such as schizophrenia). It also suggests that investigators might consider using both tasks as behavioral measures of WM in their studies to capture fully all of the aspects of WM tasks that contribute to real-world intellectual performance.

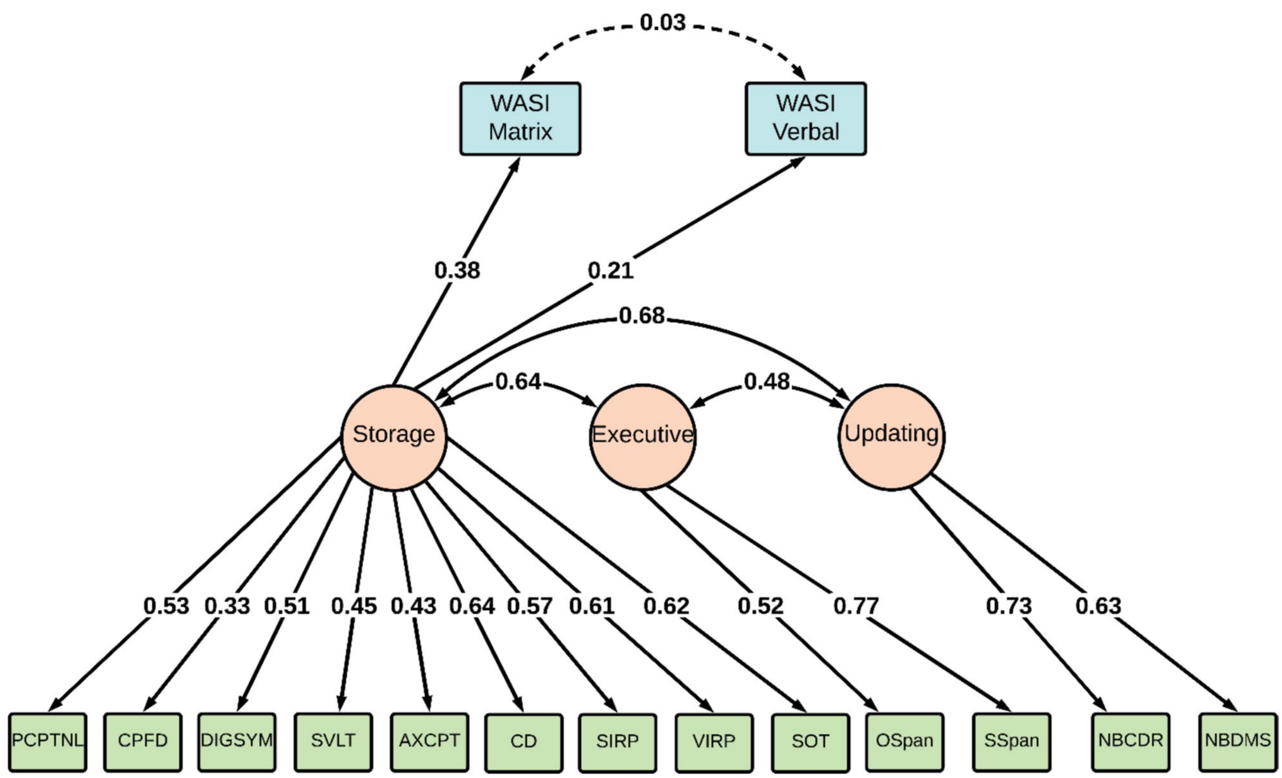


Fig. 4 The structural regression model that uses only the Storage factor to predict Wechsler Abbreviated Scale of Intelligence (2nd Edition) Vocabulary and Matrix Reasoning T Scores (WASI Verbal and Matrix). The dashed curve represents the non-significant correlation between WASI Verbal and Matrix. PCPTNL = Penn Continuous Performance Test Number Letter d'; CPFD = Penn Facial Memory Test (Delayed) d'; DIGSYM = Digit Symbol Score; SVOLTD = Short Visual Object Learning Test (Delayed) d'; AXCPT = AX Continuous Performance Test Context d'; CD = Change Detection d'; SIRP = Sternberg Item Recognition Paradigm d'; VIRP = Visual Item Recognition Paradigm d'; SOT = Self-Ordered Working Memory/pointing task Capacity Score; Ospan = Operation Span Partial Unit Score; Sspan = Symmetry Span Partial Unit Score; NBCDR = N-Back (Continuous delayed response) Accuracy; NBDMS Accuracy = N-Back (Delayed Matching-to-Sample) Accuracy. All 13 manifests were residualized measures

One particular challenge in interpreting these results is regarding the labeling and interpretation of the "Storage" factor in the current latent factor models. Because working memory tasks are not "process-pure," naming factors necessarily involves interpretations of the underlying cognitive mechanisms. In this study, tasks other than the complex span and N-back all loaded onto the same broad factor. Given their reliance on maintaining, recognizing, or responding to information, a label like Storage or Maintenance may best reflect the shared variance among these tasks, though it remains clear that multiple processes (attention, speed, long-term retrieval, short-term storage, etc.) are also involved; for example, the digit-symbol substitution task is a relatively pure measure of processing speed, while the AX-CPT is generally seen as a measure of context maintenance and cognitive control. The factor structure we observed may be predominantly driven by the unique demands of complex span tasks and N-back tasks, as compared to a specific shared feature (e.g., storage) of the other working memory and cognitive tasks included in our battery. The Executive Attention factor (two complex span tasks) reflects a more concentrated requirement to control and manipulate information in the face of concurrent processing demands from two task components, aligning with longstanding interpretations of complex span measures, while the Updating factor (two n-back tasks) emphasizes the continuous temporal embedding of stimuli, with ongoing rapid partial updating of the memory set (in the 2-back condition). Statistically, these two-task factors showed high within-factor loadings and low cross-loadings, indicating that each pair of tasks shares enough unique variance to form a coherent factor despite having only two indicators per factor. The relatively larger Storage factor, in turn, brings together a broader assortment of processes that do not cluster as neatly into smaller, more specialized constructs under an exploratory framework.

In a post hoc investigation under the CFA models, we observed that this broad Storage factor might, in principle, be split into two highly correlated subfactors, likely Primary Memory (e.g., short-term maintenance) and Secondary Memory (e.g., long-term retrieval processes). Specifically, in a more complex, confirmatory model, SVOLTD and CPFD tasks could load onto a separate secondary memory factor,

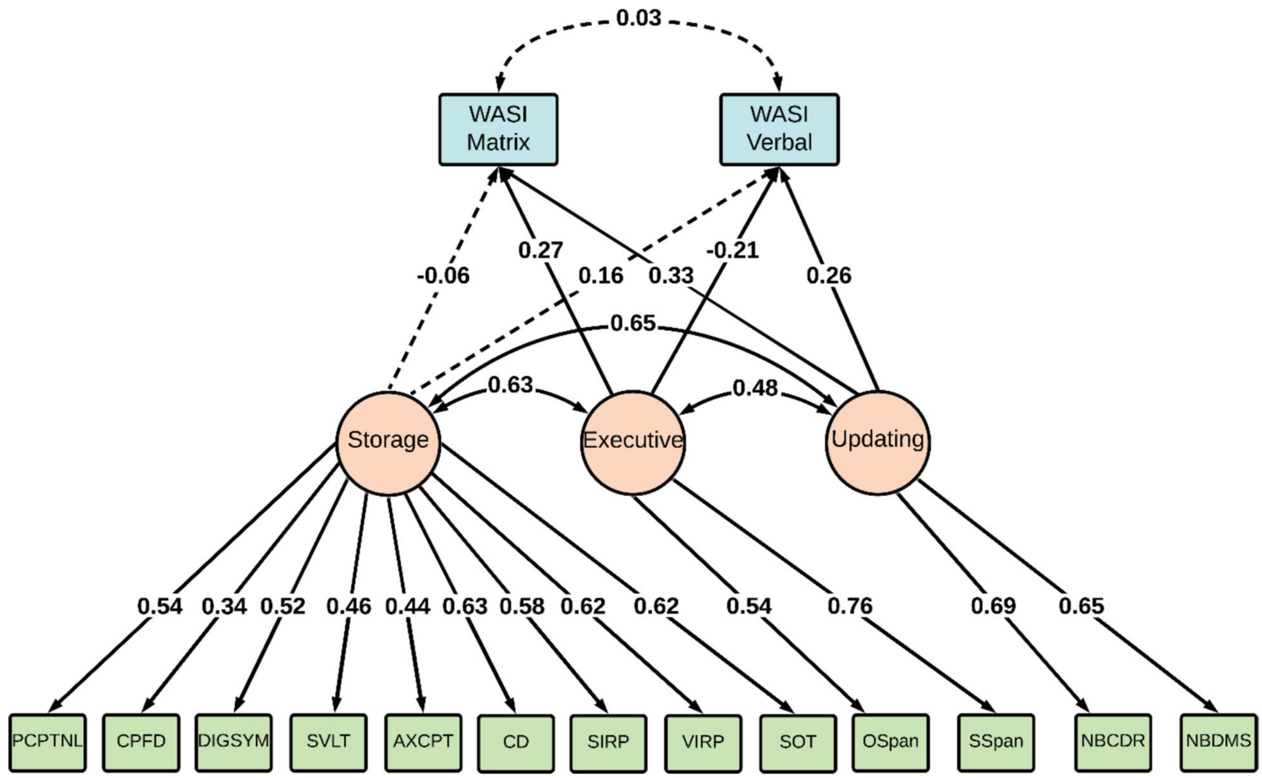


Fig. 5 The structural regression model with all three factors predicting Wechsler Abbreviated Scale of Intelligence (2nd Edition) Vocabulary and Matrix Reasoning T Scores (WASI Verbal and Matrix). The dashed curve represents the non-significant correlation between WASI Verbal and Matrix. The dashed arrows represent the non-significant structural regression coefficients of Storage to the two WASI measures. PCPTNL = Penn Continuous Performance Test Number Letter d'; CPFD = Penn Facial Memory Test (Delayed) d'; DIGSYM = Digit Symbol Score; SVOLTD = Short Visual Object Learning Test (Delayed) d'; AXCPT = AX Continuous Performance Test Context d'; CD = Change Detection d'; SIRP = Sternberg Item Recognition Paradigm d'; VIRP = Visual Item Recognition Paradigm d'; SOT = Self-Ordered Working Memory/pointing task Capacity Score; Ospan = Operation Span Partial Unit Score; Sspan = Symmetry Span Partial Unit Score; NBCDR = N-Back (Continuous delayed response) Accuracy; NBDMS Accuracy = N-Back (Delayed Matching-to-Sample) Accuracy. All 13 manifests were residualized measures

improving overall model fit compared with a confirmatory, three-factor model, $\Delta\chi^2(3) = 13.83$, $p < .01$. This is partially aligned with other findings, such as Unsworth et al. (2014) and Robison et al. (2024).

However, these two sub factors were still strongly correlated ($r = .75$), suggesting they function quite similarly in this dataset, and in the EFA they did not fully separate without each forming a single-task factor. This pattern implies that more extensive task coverage or targeted designs may be needed to produce a stable distinction between primary and secondary memory under exploratory approaches. Furthermore, based on the confirmatory measurement model with a 4-factor solution ("Primary," "Secondary," "Executive," and "Update"), a SEM model (SEM4) similar to SEM2 was estimated, in which the four correlated factors predicted the two WASI measures. In SEM4, neither Primary memory nor Secondary memory emerged as a significant predictor of the two WASI measures. Compared with previous similar research such as Robison et al. (2024), although defined by different sets of task measures, the correlation between the Primary and Secondary memory factors was higher ($r = .75$ compared with.54 in the measurement model of Robison et al.). This higher correlation was also confirmed by an additional SEM in which the two factors were specified as sperate predictors of the WASI measures. None of the regression coefficients were significant, indicating that neither primary memory nor secondary memory could explain a meaningful portion of variances in general cognitive abilities (reflected by WASI-Matrix and WASI-Verbal). From a theoretical standpoint, it suggests that although they can be distinguished, these two potential subfactors of "Storage" may overlap substantially in the current data and therefore did not explain unique variances in intelligence.

In summary, our results reinforced the idea that WM is multifaceted. Specifically, the current results align with previous research showing that complex span and n-back tasks measure different aspects of WM. The moderate correlation between the Executive Attention and the Updating factor in both the EFA and the SEM models suggests that complex span tasks and n-back tasks do

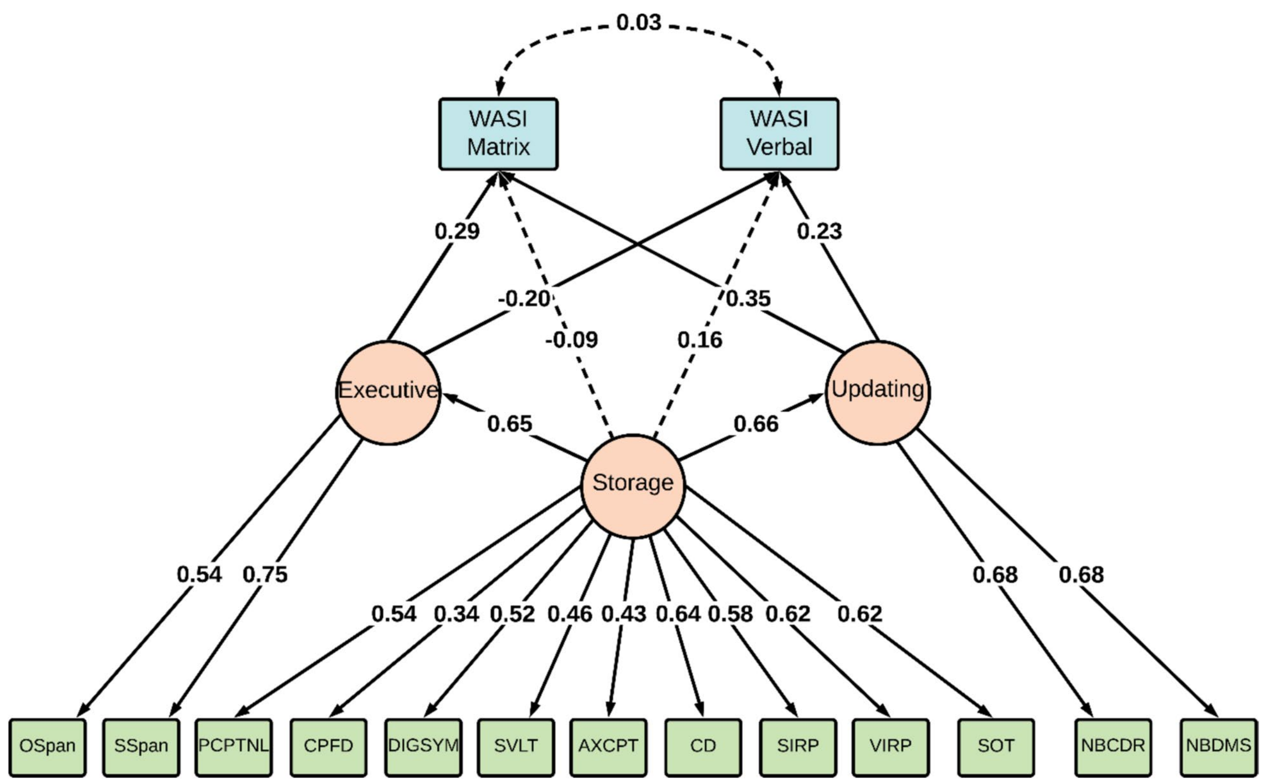


Fig. 6 The structural regression model using Executive Attention and Updating factors as mediators of the relationship between Storage factor and Wechsler Abbreviated Scale of Intelligence (2nd Edition) Vocabulary and Matrix Reasoning T Scores (WASI Verbal and Matrix). The dashed curve represents the non-significant correlation between WASI Verbal and Matrix. The dashed arrows represent the non-significant structural regression coefficients of Storage to the two WASI measures. PCPTNL = Penn Continuous Performance Test Number Letter d'; CPFD = Penn Facial Memory Test (Delayed) d'; DIGSYM = Digit Symbol Score; SVOLTD = Short Visual Object Learning Test (Delayed) d'; AXCPT = AX Continuous Performance Test Context d'; CD = Change Detection d'; SIRP = Sternberg Item Recognition Paradigm d'; VIRP = Visual Item Recognition Paradigm d'; SOT = Self-Ordered Working Memory/pointing task Capacity Score; Ospan = Operation Span Partial Unit Score; Sspan = Symmetry Span Partial Unit Score; NBCDR = N-Back (Continuous delayed response) Accuracy; NBDMS Accuracy = N-Back (Delayed Matching-to-Sample) Accuracy. All 13 manifests were residualized measures

not tap into the same set of cognitive processes, partially aligned with Kane et al. (2007) and Redick & Lindsey (2013). This distinction between Executive Attention and Updating provides further evidence for the multifaceted nature of WM, supporting theories that propose a domain-general executive control mechanism alongside domain-specific processes. Furthermore, the results from the SEM models indicated that, although the Storage factor comprising a range of common WM and other cognitive tasks was significantly associated with WASI performance, the association was no longer significant when Executive Attention and Updating were present in the model. Specifically, the relationships between Storage and the two WASI subtests were fully mediated by Executive Attention and Updating (the complex span tasks and the n-back tasks, correspondingly).

Caveats and limitations

While this study provides valuable insights into the structure of WM, several limitations should be noted. The sample consisted of undergraduate students, which may limit the generalizability of the findings to other populations. Specifically, in the current sample, it was observed in the sample that the correlation between the two WASI subtests was low (r =.11, p <.01), indicating that for the current sample of participants, their performance on verbal reasoning and performance on matrix reasoning was only weakly related. This correlation was far lower than what was reported by the WASI-II manual, where all of the subtests correlated at a least at a moderate level, ranged from 0.4 s to 0.7 s. Future research should include

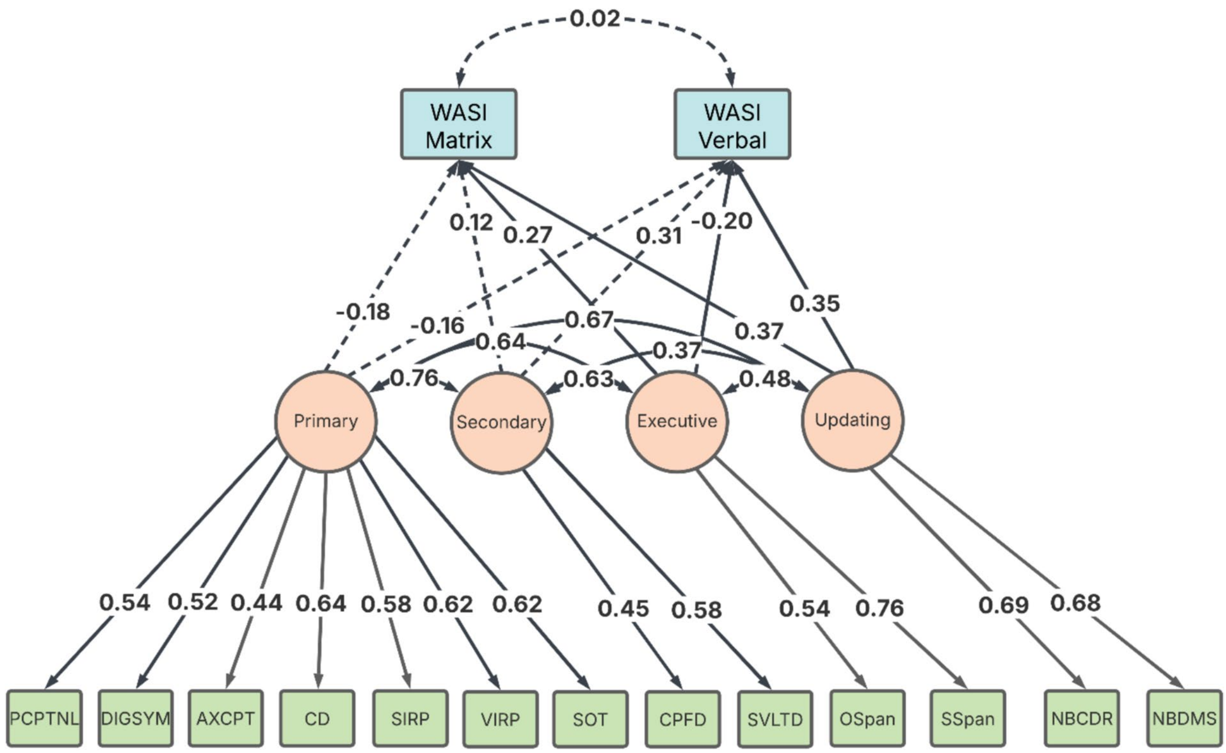


Fig. 7 The structural regression model predicting Wechsler Abbreviated Scale of Intelligence (2nd Edition) Vocabulary and Matrix Reasoning T Scores (WASI Verbal and Matrix) with "Storage" being separated into two confirmatory factors ("Primary" and "Secondary"). The dashed curve represents the non-significant correlation between WASI Verbal and Matrix. The dashed arrows represent the non-significant structural regression coefficients of Storage to the two WASI measures. PCPTNL = Penn Continuous Performance Test Number Letter d'; CPFD = Penn Facial Memory Test (Delayed) d'; DIGSYM = Digit Symbol Score; SVOLTD = Short Visual Object Learning Test (Delayed) d'; AXCPT = AX Continuous Performance Test Context d'; CD = Change Detection d'; SIRP = Sternberg Item Recognition Paradigm d'; VIRP = Visual Item Recognition Paradigm d'; SOT = Self-Ordered Working Memory/pointing task Capacity Score; Ospan = Operation Span Partial Unit Score; Sspan = Symmetry Span Partial Unit Score; NBCDR = N-Back (Continuous delayed response) Accuracy; NBDMS Accuracy = N-Back (Delayed Matching-to-Sample) Accuracy. All 13 manifests were residualized measures

more diverse samples and/or improved sampling of tasks to validate and extend these findings. Also, psychometric network modeling of the WM measures could help extend the current latent factor approach on the WM tasks and measures by relying on fewer common-cause premises to illustrate the intercorrelations (positive manifold) among the cognitive measures using partial correlation networks (Epskamp et al., 2018). Lastly, longitudinal studies could also examine how these WM components develop over time, and their stability across different contexts.

Conclusions

This study contributes to our understanding of the complex nature of WM by identifying distinct components that predict cognitive performance. The findings emphasize the importance of considering multiple WM subconstructs in research and clinical practice, and suggest that complex span and n-back tasks may be uniquely well-suited to capture aspects of WM that are related to intelligence. By recognizing the latent structure of WM based on statistically and conceptually reliable and valid measures, we can better understand individual differences in cognitive abilities and develop more effective interventions for cognitive impairments.

From a practical perspective, researchers examining WM should be aware that different tasks can emphasize distinct cognitive components. For rigorous modeling of WM's structure, it is advantageous to include multiple tasks within each theoretical subcomponent, no matter whether the goal is to partial out certain processes or identify a comprehensive metric of WM. Future studies could replicate these findings with additional tasks specifically designed to distinguish among types of storage, as well as expand the range of executive attention and updating measures.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13415-025-01310-3.

Funding Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health (NIH) under award numbers R01MH120293 to J.X.V.S. and F30MH122136 to J.C.W. J.C.W. was also supported by a Research Supplement to Promote Diversity in Health-Related Research (3R01MH120293-04S1) and by the Stony Brook University Medical Scientist Training Program (Award No. T32GM008444; Principal Investigator: Michael A. Frohman, M.D., Ph.D.). P.N.T. was supported by the Stony Brook University Scholars in Biomedical Sciences Program (NIH Award No. T32GM148331; PI: Dr. Styliani-Anna [Stella] E. Tsirka).The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Availability of data and materials The data is not immediately accessible, but it is being prepared for submission to the NIMH Data Archive (NDA) repository by February 2026 (Collection ID: C3440).

Code availability R scripts for data analyses are available at: https://osf.io/cvfta/.

Declarations

Conflicts of Interest The authors declare no conflicts of interest.

Ethics approval The study was reviewed and approved by the Institutional Review Board (IRB) at Stony Brook University.

Consent to participate Informed consent was obtained from all individual participants in the study. Participants were fully informed of the study's aims and procedures, and their right to withdraw at any time without penalty.

Consent for publication The authors consent to publication of the current study to Cognitive, Affective, and Behavioral Neuroscience.

References

Abramovitch, A., Short, T., & Schweiger, A. (2021). The C Factor: Cognitive dysfunction as a transdiagnostic dimension in psychopathology. *Clinical Psychology Review, 86*, 102007.

Baddeley, A. D. (2010). Working memory. *Current Biology, 20*, 4.

Baddeley, A. D., & Graham, J. (1974). Working memory. The Psychology of Learning and Motivation, 8, 47–89.

Baddeley, A. D., & Hitch, G. .J. (1974). Working memory. *The Psychology of Learning and Motivation, 8*, 47–89.

Barch, D. M., & Smith, E. (2008). The Cognitive Neuroscience of Working Memory: Relevance to CNTRICS and Schizophrenia. *Biological Psychiatry, 64*, 11–17.

Barch, D. M., Braver, T. S., Carter, C. S., Poldrack, R. A., & Robbins, T. W. (2009). CNTRICS final task selection: Executive control. *Schizophrenia Bulletin, 35*, 115–135.

Bowie, C. R., & Harvey, P. D. (2006). Cognitive deficits and functional outcome in schizophrenia. *Neuropsychiatric Disease and Treatment, 2*, 531–536.

Callicott, J. H., Ramsey, N. F., Tallent, K., Bertolino, A., Knable, M. B., Coppola, R., et al. (1998). Functional magnetic resonance imaging brain mapping in psychiatry: Methodological issues illustrated in a study of working memory in schizophrenia. *Neuropsychopharmacology, 18*(3), 18.

Callicott, J. H., Mattay, V. S., Bertolino, A., Finn, K., Coppola, R., Frank, J. A., et al. (1999). Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cerebral Cortex, 9*(1), 20–26.

Callicott, J. H., Bertolino, A., Mattay, V. S., Langheim, F. J. P., Duyn, J., Coppola, R., et al. (2000). Physiological dysfunction of the dorsolateral prefrontal cortex in schizophrenia revisited. *Cerebral Cortex, 10*(11), 1078–1092.

Chatham, C. H., Herd, S. A., Brant, A. M., Hazy, T. E., Miyake, A., O'Reilly, R., et al. (2011). From an Executive Network to Executive Control: A Computational Model of the n-back Task. *Journal of Cognitive Neuroscience, 23*, 3598–3619.

Chen, Y.-N., Mitra, S., & Schlaghecken, F. (2008). Sub-processes of working memory in the N-back task: An investigation using ERPs. *Clinical Neurophysiology, 119*, 1546–1559.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*, 547–552.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge University Press.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769–786.

Cowan, N. (2005). Understanding intelligence. *Handbook of understanding and measuring intelligence* (pp. 469–48). Sage Publications Inc.

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review, 26*, 197–223.

Cowan, N., Bao, C., Bishop-Chrzanowski, B. M., Costa, A. N., Greene, N. R., Guitard, D., et al. (2024). The relation between attention and memory. *Annual Review of Psychology, 75*, 183–214.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*, 422–433.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19–23.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation* (pp. 145–199). Elsevier.

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50*, 195–212.

Forbes, N. F., Carrick, L. A., McIntosh, A. M., & Lawrie, S. M. (2009). Working memory in schizophrenia: A meta-analysis. *Psychological Medicine, 39*, 889–905.

Goldman-Rakic, P. S. (1994). Working memory dysfunction in schizophrenia. *The Journal of Neuropsychiatry and Clinical Neurosciences, 6*, 348–357.

Goldman-Rakic, P. S. (1999). The physiological approach: Functional architecture of working memory and disordered cognition in schizophrenia. *Biological Psychiatry., 46*, 650–661.

Gonthier, C., Macnamara, B. N., Chow, M., Conway, A. R. A., & Braver, T. S. (2016). Inducing proactive control shifts in the AX-CPT. *Frontiers in Psychology, 7*, 1822.

Green, M. F. (1998). *Schizophrenia from a neurocognitive perspective: Probing the impenetrable darkness*. Allyn & Bacon.

Green, M. F., & Harvey, P. D. (2014). Cognition in schizophrenia: Past, present, and future. *Schizophrenia Research Cognition., 1*(1), e1–e9.

Gur, R. C., Richard, J., Hughett, P., Calkins, M. E., Macy, L., Bilker, W. B., et al. (2010). A cognitive neuroscience based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. *Journal of Neuroscience Methods, 187*, 254–262.

Henze, N., & Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods, 19*, 3595–3617.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185.

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement, 34*, 111–117.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the n-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 615–622.

Kern, R. S., Gold, J. M., Dickinson, D., Green, M. F., Nuechterlein, K. H., Baade, L. E., et al. (2011). The MCCB impairment profile for schizophrenia outpatients: Results from the MATRICS psychometric and standardization study. *Schizophrenia Research, 126*, 124–131.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Knowles, E. E. M., Weiser, M., David, A. S., Dickinson, D., Glahn, D., Gold, J., et al. (2012). Dedifferentiation and substitute strategy: Deconstructing the processing-speed impairment in schizophrenia. *Schizophrenia Research, 142*, 129–136.

Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry, 27*, 151–177.

Lee, J., & Park, S. (2005). Working Memory Impairments in Schizophrenia: A meta-analysis. *Journal of Abnormal Psychology, 114*, 599–611.

Moore, M. E., & Ross, B. M. (1963). Context Effects in Running Memory. *Psychological Reports, 12*(2), 451–465.

Morrell, R. W., & Park, D. C. (1993). The effects of age, illustrations, and task variables on the performance of procedural assembly tasks. *Psychology and Aging, 8*, 389–399.

National Institute of Mental Health. (2024). About RDoC. https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/about-rdoc

Oberauer, K. (2019). Working memory and attention – a conceptual analysis and review. *Journal of Cognition, 2*, 36.

Oberauer, K. (2019). Working memory capacity limits memory for bindings. *Journal of Cognition, 2*, 40.

Oberauer, K., & Hein, L. (2012). Attention to information in working memory. *Current Directions in Psychological Science, 21*, 164–169.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review, 19*, 779–819.

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods, 47*, 1343–1355.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*, 46–59.

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development, 14*, 187–202.

Redick, T. S., & Engle, R. W. (2011). Rapid communication: Integrating working memory capacity and context-processing views of cognitive control. *Quarterly Journal of Experimental Psychology, 64*, 1048–1055.

Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review, 20*, 1102–1113.

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., et al. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*, 164–171.

Robison, M. K., Miller, A. L., Wiemers, E. A., Ellis, D. M., Unsworth, N., Redick, T. S., et al. (2024). What makes working memory work? A multifaceted account of the predictive power of working memory capacity. *Journal of Experimental Psychology: General, 153*, 2193.

Ross, B. M. (1966). Serial order as a unique source of error in running memory. *Perceptual and Motor Skills., 23*, 195–209.

Ross, B. M. (1966). Serial-order effects in two-channel running memory. *Perceptual and Motor Skills, 23*(3), 1099–1107.

Rosseel, Y. (2012). An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

Rypma, B., Berger, J. S., & D'Esposito, M. (2002). The influence of working-memory demand and subject performance on prefrontal cortical activity. *Journal of Cognitive Neuroscience, 14*, 721–731.

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1089–1096.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*, 323–338.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*, 4–27.

Sternberg, S. (1966). High-speed scanning in human memory. *Science., 153*, 652–654.

Swanson, H. L., & Berninger, V. W. (1996). Individual differences in children's working memory and writing skill. *Journal of Experimental Child Psychology, 63*, 358–385.

Tubiolo, P. N., Williams, J. C., & Van Snellenberg, J. X. (2025). Tale of Two n-Backs: Diverging associations of dorsolateral prefrontal cortex activation with n-back task performance Journal of Neuroscience Research, 103(2), e70021.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17*, 635–654.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology, 71*, 1–26.

Van Snellenberg, J. X., Conway, A. R. A., Spicer, J., Read, C., & Smith, E. E. (2014a). Capacity estimates in working memory: Reliability and interrelationships among tasks. *Cognitive, Affective, & Behavioral Neuroscience, 14*, 106–116.

Van Snellenberg, J. X., Slifstein, M., Read, C., Weber, J., Thompson, J. L., Wager, T. D., et al. (2014b). Dynamic shifts in brain network activation during supracapacity working memory task performance. *Human Brain Mapping, 36*, 1245–1264.

Van Snellenberg, J. X., Girgis, R. R., Horga, G., Van De Giessen, E., Slifstein, M., Ojeil, N., et al. (2016). Mechanisms of working memory impairment in schizophrenia. *Biological Psychiatry, 80*, 617–626.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*, 748–751.

Wechsler, D. (2011). *Wechsler abbreviated scale of intelligence* (2nd (WASI-II)). APA PsycTests.

Open practices statement The data is not immediately accessible but is being prepared for submission to the NIMH Data Archive (NDA) repository. R scripts for data analyses are available at: https://osf.io/cvfta/. Information on data availability will be updated during the revision process. The project was not preregistered.