Published in final edited form as:
Res Dev Disabil. 2017 November 20; 72: 166–178. doi:10.1016/j.ridd.2017.10.024.

Assessing Working Memory in Children with ADHD: Minor Administration and Scoring Changes May Improve Digit Span Backward's Construct Validity

Erica L. Wells, M.S., M.Ed.[1], **Michael J. Kofler, Ph.D.**[1], **Elia F. Soto, B.S.**[1], **Hillary S. Schaefer, Ph.D.**[2,3], and **Dustin E. Sarver, Ph.D.**[4]

[1]Florida State University, Department of Psychology

[2]University of Virginia, Curry School of Education

[3]Tufts University, Institute for Applied Research in Youth Development

[4]University of Mississippi Medical Center, Department of Pediatrics, Center for Advancement of Youth

Abstract

Background—Pediatric ADHD is associated with impairments in working memory, but these deficits often go undetected when using clinic-based tests such as digit span backward.

Aims—The current study pilot-tested minor administration/scoring modifications to improve digit span backward's construct and predictive validities in a well-characterized sample of children with ADHD.

Methods and Procedures—WISC-IV digit span was modified to administer all trials (i.e., ignore discontinue rule) and count digits rather than trials correct. Traditional and modified scores were compared to a battery of criterion working memory (construct validity) and academic achievement tests (predictive validity) for 34 children with ADHD ages 8–13 (*M*=10.41; 11 girls).

Outcomes and Results—Traditional digit span backward scores failed to predict working memory or KTEA-2 achievement (all *ns*). Alternate administration/scoring of digit span backward significantly improved its associations with working memory reordering (*r*=.58), working memory dual-processing (*r*=.53), working memory updating (*r*=.28), and KTEA-2 achievement (*r*=.49).

Conclusions and Implications—Consistent with prior work, these findings urge caution when interpreting digit span performance. Minor test modifications may address test validity concerns, and should be considered in future test revisions. Digit span backward becomes a valid measure of working memory at exactly the point that testing is traditionally discontinued.

Corresponding Author: Michael J. Kofler, Ph.D., Florida State University | Department of Psychology, 1107 W. Call Street | Tallahassee, FL 32306-4301, Phone: (850) 645-0656, Fax: (850) 644-7739, kofler@psy.fsu.edu.

Conflict of Interest:
The authors have no conflicts of interest to report.

Keywords

ADHD; working memory; digit span; WISC

Working memory is a limited capacity system that involves the updating, manipulation/ serial reordering, and dual-processing of internally-held information for use in guiding behavior (Baddeley, 2007; Shelton et al., 2010; Unsworth et al., 2010). Working memory abilities have been found to directly or indirectly support myriad educational, occupational, and functional outcomes as diverse as impulse control (Raiker et al., 2012), delay tolerance (Patros et al., 2015), and skill in strategy-based games such as chess and bridge (Baddeley, 2007). In addition, impaired working memory has been implicated in hyperactive behavior (Rapport et al., 2009), inattention (Kofler et al., 2010), and mind wandering (Kane et al., 2007), and may be etiologically important for a broad range of psychopathologies including schizophrenia (Goldman-Rakic, 1994), depression (Joormann & Gotlib, 2008), and ADHD (Kasper et al., 2012).

Recent critiques of the clinical literature, however, raise questions about conclusions regarding working memory's association with psychopathology (Snyder et al., 2015). In particular, converging evidence questions the construct validity (Redick & Lindsey, 2013) and structural organization (Friedman & Miyake, 2016) of common neuropsychological tests used to assess working memory. Moreover, the clinical literature has been criticized for using executive functioning tests that are not cognitively informed according to contemporary advancements (Snyder et al., 2015). The evidence base for these critiques is substantial, and includes data from large samples of healthy and clinically referred children, adolescents, and adults (Bowden et al., 2013; Colom et al., 2005; Egeland, 2015; Perry et al., 2001; Tarle et al., 2017; Twamley et al., 2006). The current study's goal was to begin to address these criticisms by piloting relatively minor and straightforward test modifications involving scoring, task demands, and administration parameters to improve our ability to capture construct-relevant individual differences in working memory. To do so, we selected a working memory test commonly used in clinical practice (digit span backward; Martinussen et al., 2005), and compared performance among children with ADHD based on traditional vs. alternate administration/scoring to a battery of criterion working memory (construct validity) and academic achievement tests (predictive validity). Selecting a pediatric ADHD sample is ideal for testing these effects because disrupted working memory may be implicated in the disorder's developmental pathways (Halperin et al., 2008; Kasper et al., 2012; Rapport et al., 2013), and because ADHD is common disorder presenting for clinical assessment (Rowland et al., 2002).

Working memory vs. short-term memory

Clarifying the extent to which a test measures working memory, rather than short-term memory, is critical given that these abilities are associated with anatomically distinct cortical regions (Nee et al., 2013; Wager & Smith, 2003), support functionally distinct cognitive processes (Alloway et al., 2006), and show unique relations with important academic and functional outcomes (Rapport et al., 2013). Short-term memory involves the temporary storage and rehearsal of information, and reflects the *memory* component of working memory (Baddeley, 2007; Shelton et al., 2010; Unsworth et al., 2010). In contrast, *working* memory refers to the active, top-down manipulation of information held in short-term memory, and includes interrelated functions of the mid-lateral prefrontal cortex that involve continuous *updating* (active addition and deletion of items from working memory), *dual-processing* (diverse processes that involve operating on information while storing the same or other information in working memory), and *serial reordering* (mental manipulation of temporal/sequential order) (Nee et al., 2013; Wager & Smith, 2003).

Relevant to clinical/educational assessment in general and pediatric ADHD in particular, working memory underlies learning skills such as math and reading attainment (Raghubar et al., 2010; Sesma et al., 2009; Swanson & Kim, 2007), science (Gathercole et al., 2004), written language (Alloway et al., 2005), oral language (McInnes et al., 2003), and following directions (Jaroslawska et al., 2016). In contrast, short-term memory usually demonstrates weaker relations with reading and math achievement, and is generally unrelated to comprehension and complex learning (Rapport et al., 2013).

Is digit span backward an optimal measure of working memory?

Digit span is one of the most commonly used tests of working memory in clinical research and practice (Kasper et al., 2012). Most digit span tests contain two, sequentially presented parts: digit span forward and digit span backward. Digit span forward involves the immediate, rote recall of orally presented numerical sequences and is considered a measure of short-term memory. Digit span backward involves similar, orally presented numerical sequences, but is interpreted as a measure of *working* memory because it requires respondents to recall digits in reverse serial order (e.g., 5-2-7 is correctly recalled as 7-2-5). This reversal demand solely differentiates the two digit span conditions; thus, the extent to which list reversal evokes sufficient updating, reordering, and/or dual-processing to engage working memory is critical to its clinical and theoretical utility (Wechsler, 2003).

Beyond the face validity of list reversal as an active reordering process, evidence supporting the use of digit span backward as a measure of working memory includes findings that children, adolescents, and adults tend to recall fewer trials correctly and recite shorter maximum lists of numbers on digits backward relative to digits forward tasks (Conklin et al., 2000; Gathercole et al., 2004; St. Clair-Thompson, 2010). In addition, these forward and backward simple span tasks show convergent validity with similar, rote recall and list reversal tests (Wechsler, 2003). At the same time, digit span backward's reversal demands may not be sufficient to evoke the dual-processing demands characteristic of the working memory construct (Conway et al., 2005). That is, evidence suggests that digit span backward *and* digit span forward load on a single short-term memory factor, and load separately from well-validated working memory complex span tasks (Ackerman et al., 2005; Engle et al., 1999) in healthy adults (Colom et al., 2005), healthy children (Swanson & Kim, 2007), adults with psychopathology (Perry et al., 2001; Twamley et al., 2006), and adolescents/ adults referred for neurological evaluation (Bowden et al., 2013; Egeland, 2015). In addition, these short-term and working memory factors predict non-overlapping variance in outcomes such as IQ (Unsworth & Engle, 2007). Similarly, and most relevant to the current study, two recent studies reported that WISC-IV digit span backward fails to correctly classify ADHD and Non-ADHD children above chance levels (Colbert & Bo, 2017; Tarle et al., 2017), whereas criterion working memory tasks demonstrate group classification accuracy at levels similar to or better than gold standard ADHD rating scales (Tarle et al., 2017). Taken together, these findings provide strong evidence that, “a simple transformation of order [from forward to backward] would be insufficient to move a task from the short-term memory storage category to the working memory category” (Engle et al., 1999, p. 314).

Can minor administration/scoring changes improve digit span backward’s construct validity?

The poor correspondence between digit span backward and working memory processes may be due to digit span’s administration conventions. In the Wechsler (2003) digit span backward test, for example, list-length increases by one digit every two trials, and testing is discontinued as soon as the child demonstrates less than perfect recall of both same-length trials. This approach has been criticized for providing a relatively narrow range of scores (Tarle et al., 2017; Woods et al., 2011), such that score variability comes primarily from the shortest list-lengths because longer list-lengths are not routinely administered (Conway et al., 2005). These shorter list-lengths, in turn, may not exceed short-term memory capacity and thus fail to engage sufficient working memory processes in the absence of additional processing demands (Unsworth & Engle, 2006).

In addition, most digit span tasks use all-or-nothing scoring, which has been criticized psychometrically and shows significantly reduced reliability relative to partial-credit scoring (Conway et al., 2005). All-or-nothing scoring assigns credit only for perfectly recalled sequences, whereas partial-credit scoring counts each digit recalled in the correct serial position. Like the discontinue rules above, this process blunts scoring variability and thus decreases the test’s ability to detect individual differences (Tarle et al., 2017). To test these hypotheses, Unsworth & Engle (2006) administered all trials of a simple span task (i.e., omitted the task’s discontinue criteria) and scored each stimulus recalled in the correct serial position (partial-credit scoring). They found that simple span predicts complex span (working memory) in healthy adults only at simple span list-lengths that exceed short-term memory capacity. The current study adapted this methodology to determine whether these high list-length trials show improved correspondence with working memory in a clinical child sample, and whether this improved correspondence was attributable to partial-credit scoring, administering trials beyond the traditional discontinue rule, or both.

Present Study

Taken together, this literature suggests a conflict between interpretation of simple span tasks, particularly digit span backward, as indicators of *working* memory, and highlights the scoring, task demands, and administration parameters considered necessary to sufficiently evoke working memory processes. Previous findings from healthy adult samples suggest that relatively minor and straightforward modifications may allow digit span tasks to optimally estimate working memory. To our knowledge these issues have never been examined in a clinical child sample. Further, comparisons between clinical tests and criterion working memory measures have generally been limited to simple vs. complex span tasks despite meta-analytic evidence for three interrelated types of working memory (dual-processing, updating, serial reordering; Wager & Smith, 2003). It seems reasonable that digit span backward's reversal demands may show stronger continuity with serial reordering relative to other core working memory processes.

The current study recruited a well-defined, pediatric ADHD sample and used the Unsworth and Engle (2006) modifications to assess the construct and predictive validities of digit span backward (derived from WISC-IV) relative to criterion measures of working memory, academic achievement, and fluid intelligence (Conway et al., 2005). We hypothesized that digit span backward would correlate moderately with IQ but show lower associations with working memory and academic achievement (i.e., poor construct and predictive validity, respectively). We expected the alternate administration/scoring rules to significantly improve digit span backward's construct and predictive validities, with these improvements coming specifically from performance on trials that are not routinely administered due to the test's discontinue rule.

Method

Participants

Thirty-seven children with ADHD participated from 2013–2014; recruitment was closed due to the principal investigator's change of institution. Digit span data were missing for three children due to experimenter error (following discontinue rule), resulting in a final *N* of 34 children, aged 8 to 13 years of age (*M* = 10.41, *SD* = 1.38; 23 boys, 11 girls) from the Southeastern United States. Children were consecutive referrals to a university-based children's learning clinic (CLC) through community resources to participate in a larger study examining predictors of behavioral treatment response for ADHD. Pre-treatment data were used. Performance on the working memory reordering tests was reported in [REMOVED FOR BLIND REVIEW] to investigate conceptually unrelated hypotheses. Psychoeducational evaluations were provided to the parents of all participants. All parents and children gave informed consent/assent; Institutional Review Board approval was obtained prior to data collection. Sample ethnicity was representative of the region: Caucasian non-Hispanic (80%), mixed racial/ethnic (11%), Hispanic English-speaking (6%), and Asian (3%).

Diagnostic evaluation—All children and their parents participated in a detailed, semi-structured clinical interview using the Kiddie Schedule for Affective Disorders and Schizophrenia for School-Aged Children (K-SADS; Kaufman et al., 1997). The K-SADS (2013 Update) assesses onset, course, duration, severity, and impairment of current and past episodes of psychopathology in children based on DSM-5 criteria, and was supplemented with parent and teacher ratings from the Behavior Assessment System for Children (BASC-2; Reynolds & Kamphaus, 2004) and Child Symptom Inventory (CSI-IV; Gadow & Sprafkin, 2002). Children were excluded for gross neurological, sensory, or motor impairment; seizure disorder, psychosis, autism spectrum, or intellectual disability; or non-stimulant medications that could not be withheld for testing.

Inclusion criteria included: (1) DSM-5 ADHD diagnosis by the directing clinical psychologist based on K-SADS; (2) parent ratings ≥1.5 *SD*s on the BASC-2 Attention Problems and/or Hyperactivity scales, or exceeding criterion scores on the parent CSI-IV ADHD-Inattentive and/or ADHD-Hyperactive/Impulsive subscales; and (3) teacher ratings ≥1.5 *SD*s on the BASC-2 Attention Problems and/or Hyperactivity scales, or exceeding criterion scores on the teacher CSI-IV ADHD-Inattentive and/or ADHD-Hyperactive/Impulsive subscales. All children had current impairment based on K-SADS. Four participants failed to meet teacher criteria, likely due to behavior well controlled by medication, but were included based on previous psychoeducational evaluations that documented cross-setting symptoms/impairment.

Of the 34 children with ADHD, 14 met criteria for Combined, 19 for Inattentive, and 1 for Hyperactive/Impulsive Presentation. To improve generalizability, children with comorbidities were included. Comorbidities reflect clinical consensus best estimates, and included oppositional defiant (14%), depressive (16%), and anxiety disorders (16%). Sixteen of the 34 children with ADHD were currently prescribed psychostimulants. Medication was withheld for a minimum of 24 h prior to both research sessions.

Procedures

All children participated in two, 3-hour research sessions following the psychoeducational assessment. The working memory tests were administered as part of a larger battery of counterbalanced laboratory tasks. All children received brief (2–3 min) breaks following each task, and preset longer (10–15 min) breaks after every two to three tasks to minimize fatigue.

Neurocognitive Performance

Digit Span—The Digit Span subtest from the WISC-IV (Wechsler, 2003) was administered with the modifications described by Unsworth and Engle (2006). Standardized WISC-IV administration was followed with the exception of the discontinue rule: Children received all trials of both digit span forward (DSF; 2 trials each at list lengths 2–9) and digit span backward (DSB; 4 trials at list length 2, 2 trials each at list lengths 3–8). Internal consistency reliability in the current sample was $\alpha = .85$ (DSF) and $\alpha = .73$ (DSB).

Working memory reordering—The phonological and visuospatial working memory reordering tasks developed by Rapport et al. (2009) were used for the current study. These tasks demonstrate large magnitude performance differences between ADHD and typically developing children (Patros et al., 2015), and performance predicts ADHD-related impairments in objectively-measured activity level (Rapport et al., 2009), attentive behavior (Kofler et al., 2010), and impulsivity (Raiker et al., 2012). Reliability evidence includes high internal consistency (α=.82–.97) and 1–3 week test-retest reliability (*r*=.76–.90; Sarver et al., 2015). Internal consistency in the current sample was α=.95 (PHWM) and α=.96 (VSWM).

Both tasks involve serial reordering of characters presented (numbers, black dot locations), and reordering of a target stimulus (letter, red dot location) into the final serial position recalled. The target stimulus was never presented in the first or last position of the sequence to minimize primacy and recency effects, and was counterbalanced across trials to appear an equal number of times in the other serial positions (i.e., position 2, 3, 4, or 5). Trials were presented sequentially in 12-trial blocks of increasing set size, from 3 to 6 stimuli per trial (48 total trials per task), with short breaks between each block (approximately 1 minute). Five practice trials were administered before each task (80% correct required). Task duration was approximately 2.5 (visuospatial) to 3.5 (phonological) minutes per block. Partial-credit scoring was used at each set size as recommended (Conway et al., 2005).

Phonological (PH) working memory reordering task: The phonological working memory task is similar to the Letter-Number Sequencing subtest on the WISC-IV (Wechsler, 2003), and assesses phonological working memory based on Baddeley's (2007) model. Prerecorded stimuli were presented aurally (1000 ms presentation + ISI) to isolate phonological task demands by minimizing visuospatial influences (Alderson et al., 2015). Children were instructed to recall the numbers in order from least to greatest, and to say the letter last (e.g., 4-H-6-2 is correctly recalled as 2-4-6-H). Two trained research assistants, shielded from participant view, recorded responses independently (interrater reliability=99.50%).

Visuospatial (VS) working memory reordering task: Children were shown 9 squares arranged in three offset vertical columns on a computer monitor. The columns were offset from a standard 3×3 grid to minimize the likelihood of phonological coding of the stimuli. A series of 2.5 cm diameter dots (3, 4, 5, or 6) were presented sequentially in one of the nine squares during each trial such that no two dots appeared in the same square on a given trial. Each dot was displayed for 800 ms followed by a 200 ms ISI. Children were instructed to input the spatial locations of the black dots in the order presented followed by the red dot's position last on a modified keyboard.

Working memory dual-processing—Working memory complex span tasks interleave the presentation of to-be-remembered target stimuli, such as digits or words, with a demanding, secondary processing task, such as verifying equations or counting shapes (Conway et al., 2005). Engaging in a secondary task of the same modality as the target stimuli yields interference effects that increase demands on controlled attention and the central executive, because both processes rely on the same limited-capacity phonological store (Conway et al., 2005).

The counting span and operation span working memory tasks described by Conway et al. (2005) were adapted for use with children. Both assess dual-processing working memory based on the Engle et al. (1999) model. Comparisons of ADHD and typically developing children indicate medium to large magnitude between-group differences on similar complex span tasks (Kuntsi et al., 2001; Willcutt et al., 2001). Evidence for reliability and validity of working memory complex span tasks includes high internal consistency (α=.77–.81), 3-month test-retest reliability (*r*=.70–.80), and expected relations with other measures of working memory (Conway et al., 2005). Eight total trials (2 per set size, presented sequentially) were completed for each complex span task, with two practice trials administered prior to advancing to the full task (100% correct required). The secondary processing task was either experimenter-paced (counting span) or self-paced (operation span) by design to allow removal of processing speed demands when estimating shared variance as described below. Serial position was a criterion for correct responses for counting span but not operation span based on the same rationale. Task duration was approximately 3.5 (Counting Span) and 5.5 (Operation Span) minutes. Internal consistency reliability in the current sample was α=.77 (Counting Span) and α=.81 (Operation Span).

Counting span: Children were sequentially shown screens containing a random number of black dots and between 1 and 9 red dots (all 2.5 cm diameter). Children were instructed to verbally report the number of red dots as each screen was presented, ignoring the black dots. After a predetermined number of screens (set sizes 3, 4, 5, and 6), children were asked to recall the number of red dots on each screen in serial order. Each screen was displayed for 500 ms per red dot (e.g., screens with 6 red dots remained visible for 3000 ms; ISI = 500 ms).

Operation span: The operation span task required children to mentally hold a sequence of unrelated, monosyllabic words in mind while simultaneously evaluating simple math equations (e.g., 1 + 3 = 2, Yes/No). A new, to-be-recalled word is presented for 1 second after each math problem, followed by a 500 ms interstimulus interval. Children were instructed to read each math problem aloud, verbally respond “yes” or “no,” and then say the to-be-recalled word. After a predetermined number of simple math problems (set sizes 3–6), children were asked to recall the words; credit was earned for each correctly recalled word regardless of serial position.

Working memory updating—Working memory updating tasks involve the constant monitoring and rapid addition/deletion of working memory contents (Miyake & Friedman, 2012). The N-back task is arguably the most commonly used continuous updating test and shows acceptable convergence with conceptually distinct measures of working memory, including complex span and serial reordering tasks (Schmiedek et al., 2014). The high-density, double-letter (1-back) N-back task described by Denney et al. (2005) was used in the current study. Comparisons of ADHD and typically developing children indicate large magnitude group differences on this task (Raiker et al., 2012). Evidence for reliability and validity includes high internal consistency (r_{block}=.66–.90) and demonstration of the expected magnitude of relations with a 0-back version of the test (Denney et al., 2005). A practice block of 30 stimuli (10 targets) was included; children were required to achieve 80% correct before advancing to the full task. Internal consistency reliability for the 3 blocks in the current sample was α=.94.

N-back: The N-back task displayed a total of 540 capital letters (3.5 cm height and width), one at a time (200 ms presentation, 800 ms ISI). Children were instructed to press a mouse button each time a target letter was identical to the letter immediately preceding it (i.e., 1 back in the sequence). One-back targets comprised 180 (33.3%) of the 540 stimuli (Denney et al., 2005). Children completed three consecutive blocks of 180 stimuli (60 targets per block); task duration was nine minutes. Total errors per block served as the primary indices of working memory updating, computed as omissions (missed targets) plus commissions (incorrect identifications; Denney et al., 2005).

Dependent variables: Working memory reordering, dual-processing, & updating: Estimates of the three working memory subprocesses were computed separately by combining the multiple blocks/set sizes per task using the dimension reduction approach described by Rapport et al. (2009). Statistically, this involves computing a Bartlett weighted average based on the intercorrelations among task performance scores (DiStefano et al., 2009). Conceptually, this process isolates "common and perfectly reliable variance" (Swanson & Kim 2007, p. 158) associated with the *working* components of working memory by removing task-specific demands associated with non-executive processes and the anatomically and functionally distinct short-term memory subsystems (Alloway et al., 2006; Fassbender & Schweitzer, 2006; Wager & Smith, 2003). The ratio of participants (34) to components (1) was deemed acceptable for each of the three computations (Hogarty et al., 2005). *Working memory reordering* was estimated as the weighted average of the eight phonological and visuospatial reordering scores (set sizes 3–6 per task; 82.55% variance accounted; loadings=.82–.96; eigenvalue=3.30). *Working memory dual-processing* was estimated similarly from the eight operation span and counting span performance scores (set sizes 3–6 per task; 73.42% variance accounted; loadings=.79–.90; eigenvalue=2.94). *Working memory updating* was estimated similarly from the three n-back blocks (89.77% variance accounted; loadings=.94–.97; eigenvalue=2.69).

Evidence for the reliability, construct validity, and predictive validity of the working memory variables includes strong internal consistencies (.77–.96), intercorrelations with each other (.49–.69), and associations with academic achievement (.43–.73) in the current sample. As expected (Conway et al., 2005), working memory reordering correlated strongly with dual-processing (r=.69) and updating (r=.61); the dual-processing/updating relation was acceptable (r=.49).

Academic Achievement

Kaufman Tests of Educational Achievement (KTEA-2)—The KTEA-2 *Comprehensive Academic Composite* (Kaufman & Kaufman, 2004) was used to assess academic achievement.

Socioeconomic Status and Measured Intelligence

Socioeconomic status—Socioeconomic status (SES) was estimated using the Hollingshead (1975) scoring based on caregiver(s)' education and occupation.

WASI-II IQ—All children were administered the Perceptual Reasoning Index (PRI; Matrix Reasoning, Block Design) from the WASI-II (Wechsler, 2011). Following Rapport et al. (2009), a residual PRI score was derived by covarying the working memory variables described above out of PRI (R^2=.27, p=.02). This residual PRI score represents cognitive functions important for IQ test performance but unrelated to working memory, and was computed to improve construct specificity because PRI performance depends heavily on working memory, even for IQ subtests not reified as 'working memory' measures (Ackerman et al., 2005; Dennis et al., 2009). An expanded discussion of the rationale and support for this method is provided in Kofler et al. (2016).

Data Analysis Overview

Demographic and test performance variables are shown in Table 1. We first replicated previous reports regarding digit span backward's construct validity, and then tested the extent to which our administration/scoring modifications improved its association with working memory (Tier 1). Digit span backward's predictive validity was then assessed relative to KTEA-2 achievement (Tier 2). Statistical significance was determined using 95% confidence intervals constructed using bias-corrected bootstrapping procedures (k=5000 bootstrap samples). Confidence intervals that did not include 0.0 were considered significant.

As shown in Table 2 (bold cells), we scored digit span backward according to the WISC-IV manual (all-or-nothing scoring of trials at/below the discontinue rule; Wechsler, 2003). We then computed partial-credit scores (total stimuli correct) for high list-length trials that are not routinely administered according to the standardized discontinue rule (Table 2, right). We defined high list-length trials based on the dichotomy used for Unsworth and Engle's (2006) adult sample (low[1]: 2–4 digits vs. high: 6–8 digits) given evidence that short-term memory and working memory are differentiated in children as young as 6 (Gathercole et al., 2004) and that short-term memory reaches an adult-like 4±1 chunk capacity during the current study's age range (Cowan et al., 2001, 2010; Tillman et al., 2011).

Tests of dependent *r*

Tests of dependent r (zr_{dep}; Lee & Preacher, 2013) examined the extent to which these alternate administration/scoring methods significantly improved digit span backward's construct validity relative to traditional administration/scoring. P-values are reported for tests of dependent r.

Results

Power Analysis

Given the relatively small sample size, we conducted a power analysis using GPower (v3.1; Faul et al., 2007) to determine our sensitivity for detecting effects. Results indicated that we are powered to detect associations of $r \geq .40$ between digit span backward and the criterion working memory variables based on our final sample size of 34 (power=.80, α=.05). Effects of this magnitude were considered reasonable given the hypothesis that these variables measure the same construct (Conway et al., 2005). We are unaware of any rules of thumb regarding the minimum correlation needed to conclude that two variables index the same construct. However, for illustrative purposes, we note recent reviews in the working memory literature have cited $r \approx .60$ as evidence of a single underlying construct (Conway et al., 2005, p. 779). The current study is powered > .99 to detect effects of this magnitude. Conversely, a recent review used $r \approx .20$ as evidence against using a task to measure working memory (Redick & Lindsey, 2013).

[1]Analysis of low list-length trials produced results that were highly similar to those reported here based on traditional administration and scoring. Similarly, the pattern and interpretation of results was unchanged when defining high list-lengths individually based on each child's performance (i.e., trials beyond the traditional discontinue rule) rather than based on theory and precedent as reported herein (Unsworth & Engle, 2006). These analyses are shown in Supplementary Tables 2–4.

For comparative purposes, intercorrelations among our working memory reordering, dual-processing, and updating variables were all significant and ranged from .49 to .69 (95% CIs exclude 0.0), suggesting sufficient power to evaluate construct validity.

Preliminary Analyses

No univariate/multivariate outliers were identified. Medication status, ADHD presentation, and gender were not significantly correlated with any variables of interest (all 95%CI include 0.0) and therefore not included as covariates. Age was associated with digit span backward performance at high list-lengths (*r*=.37). SES was associated with working memory dual-processing (*r*=.35) (95%CIs exclude 0.0). Therefore, the analyses reported below reflect partial correlations corrected for age and SES.

Tier 1. Construct validity of digit span backward

All-or-nothing (traditional) scoring—As expected, digit span backward was more difficult than digit span forward (Table 1), as evidenced by lower raw scores (paired-sample *t*[33]=2.20, 95%CI=1.61–2.84) and span scores (*t*[33]=2.09, 95%CI=1.63–2.53). However, this difficulty was not attributable to increased working memory demands because digit span backward *was not related* significantly to working memory reordering (*r*=.03), dual-processing (*r*=.22), or updating (*r*= −.03) (all 95%CIs include 0.0; Table 2, left). In contrast, digit span backward showed strong association with fluid IQ (*r*=.49, 95%CI excludes 0.0).

Partial-credit scoring of high list-length trials—As hypothesized, the combination of high list-length administration and partial-credit scoring produced moderate-to-strong, significant associations with working memory reordering (*r*=.58), dual-processing (*r*=.53), and updating (*r*=.28) (95%CIs exclude 0.0). Tests of dependent *r* revealed that these relations were significantly higher than those obtained from traditional administration/scoring for reordering (zr_{dep} = 4.38, $p < .0005$), dual-processing (zr_{dep} = 2.44, $p = .01$), and updating (zr_{dep} = 2.22, $p = .03$).

Tier 2. Predictive validity of digit span backward

As shown in Table 3, traditionally administered/scored digit span backward failed to predict KTEA-2 academic achievement (*r*=.26, *ns*). This relation became significant, however, based on partial credit scoring of high list-length trials (*r*=.49, 95%CI excludes 0.0)(zr_{dep} = 1.78, $p = .07$). For comparison, the three criterion working memory variables showed strong predictive validity for KTEA-2 achievement (*r*=.43–.73; 95%CIs exclude 0.0; Table 3).

Tier 3. Sensitivity analyses

Finally, we conducted exploratory analyses to determine whether partial credit scoring alone was sufficient to produce the improved construct and predictive validities reported above. This involved computing scores based on partial-credit scoring and traditional administration (equivalent to scores that would be obtained by rescoring existing digit span backward datasets that stopped administration based on the discontinue rule). Unfortunately, scores based on this method alone failed to predict dual-processing (*r*=.20), updating (*r*= −.05), reordering (*r*=.11), or academic achievement (*r*=.20)(all *ns*: Supplementary Tables 2–4).

Tier 4. Digit span forward vs. backward

Although not the study's primary focus, exploratory analyses were conducted using digit span forward to provide initial insights into the role of digit span backward's list reversal demands. Results are shown in the Supplementary Materials (online), and in combination with the results above indicate that improved construct and predictive validities require (1) the administration of high list-length trials, (2) partial-credit scoring, *and* (3) list reversal demands. That is, both forward and backward digit span showed moderate-to-strong association with working memory based on partial-credit scoring of high list-length trials; however, with these modifications only digit span backward showed the expected correspondence with norm-referenced academic achievement tests.

Discussion

The current study was the first to pilot modified administration and scoring methods to address contemporary criticisms regarding digit span backward's construct and predictive validities. Our criterion working memory variables showed strong internal consistency (α=.77–.96), strong interrelations with each other (*r*=.49–.69), and expected associations with academic achievement (*r*=.43–.73) in the current sample. This strong reliability, construct validity, and predictive validity evidence suggests confidence in their use as criterion tests to which digit span backward is compared.

Overall, results replicated consistent findings from healthy and clinical child and adult samples (Egeland, 2015; Engle et al., 1999; Swanson & Kim, 2007; Tarle et al., 2017), and provided additional evidence that digit span backward does not capture individual differences in various working memory processes when administered and scored using standardized procedures. However, our pilot study produced evidence that these nonsignificant associations could be substantially enhanced to provide reasonable-to-strong estimates of several core working memory processes. This pattern indicates that small changes in administration and scoring can evince large alterations in our ability to capture theoretically meaningful cognitive processes.

As expected, we found that digit span backward was more difficult than digit span forward, as evidenced by significantly fewer correct trials and lower simple spans in our small but adequately powered sample. The consistency of our findings with those from larger ADHD samples (Kasper et al., 2012; Martinussen et al., 2005; Tarle et al., 2017), the test's standardization sample (Wechsler, 2003), and healthy child samples (Monaco et al., 2013; St. Clair-Thompson & Allen, 2013) suggests confidence in their generalizability and confirms that list reversal is more cognitively demanding than rote recall (St. Clair-Thompson & Allen, 2013).

This increased difficulty has been interpreted as evidence that list reversal evokes sufficient processing demands to engage working memory, and backward digit span is described as a test of working memory across a broad range of commonly used clinical and neuropsychological test batteries. The evidence base is at odds with this interpretation, and the current findings provide additional evidence of minimal involvement of working memory in digit span backward performance (r_{range}= −.03 to .22). At the same time, digit span backward's association with IQ was substantial (*r*=.49), suggesting that its increased difficulty may be attributable, in part, to processes involved in fluid reasoning rather than working memory.

Of primary interest in the current study was the extent to which easy-to-implement changes in administration (eliminating the discontinue rule) and scoring (assigning partial credit) could address criticisms regarding digit span backward's construct validity. Together, these modifications significantly improved digit span backward's construct validity relative to criterion tests of working memory reordering and dual-processing (*r*=.53–.58). Importantly, this improvement was attributable to performance during trials that would not have been administered in routine clinical practice due to the test's discontinue rule. That is, digit span backward became a strong approximation of working memory processes only during trials that exceeded short-term memory capacity (Unsworth & Engle, 2006).

Similarly, performance at these high list-length trials produced strong predictive validity evidence relative to KTEA-2 achievement (*r*=.49). This finding was similar to the working memory/achievement associations obtained for our criterion working memory variables (*r*=.43–.73), and is highly consistent with findings from healthy samples (e.g., Swanson & Kim, 2007). Taken together, these findings provide initial evidence that modifying digit span backward may better capture construct-relevant individual differences in a clinical child sample for whom working memory may be etiologically important (Kasper et al., 2012).

List reversal vs. list reordering: Why does only list reordering engage working memory?

It seemed reasonable to expect that traditionally-scored digit span backward would show higher correspondence with working memory reordering than dual-processing, given the face valid similarities between stimuli reversal and reordering. However, this was not the case. The digit span backward/reordering association was not appreciably different from zero (*r*=.03), whereas the non-significant digit span backward/dual-processing association (*r*=.22) may have reached significance in a much larger sample ($N \geq 125$ based on power analysis). This pattern of results may be related to our use of reordering tasks that differed by stimulus modality (phonological, visuospatial), relative to dual-processing tasks that both rely on phonological memory. Thus, the modest digit span backward/dual-processing correlation may be attributable to shared verbal short-term storage processes (Baddeley, 2007) rather than the construct of interest (working memory).

The distinction between reversal and reordering tasks is not readily apparent despite current and previous evidence that these tasks show minimal correspondence (Conway et al., 2005). A potential explanation for this discrepancy may be the higher interference effects caused by stimuli during reordering relative to reversal tasks (Conway et al., 2005). That is, reordering digits from least to greatest (in the paradigm we used) requires access to long-term memory to retrieve previously learned numerical order and update the list as each digit is re-sequenced, which may interfere with concurrent maintenance of the original memory set. As such, we hypothesize that reversal tasks like digit span backward may not require the assumed mental manipulation processes but rather increased short-term rehearsal load. In this case, children may not be mentally reversing the digits to create a new reverse-ordered list, but rather rehearsing the original list more than once and saying the next-to-last value in the sequence each time.

Alternatively, our informal conversations with children suggest that some may rely on mental visualization processes to complete the task (i.e., visualizing the sequence, then reading it backwards). This latter observation is consistent with the current finding that digit span backward, a verbal task, correlated highly with non-verbal fluid reasoning (r=.49), as well as evidence that digit span backward cross-loads with visual-spatial memory tasks (Jaroslawska et al., 2016). In either case, the poor correspondence between digit span backward and working memory reordering may be because the former can be completed without the specific mental manipulation processes that characterize the working memory construct (Baddeley, 2007).

Implications for core processes and ADHD

The present results suggest that digit span backward's task parameters may prevent most children with ADHD from reaching the higher list-length trials that engage working memory by exceeding short-term storage capacity (Unsworth & Engle, 2006). Exceeding short-term memory capacity, in turn, has been shown to evoke higher levels of inattentive behavior in both ADHD and non-ADHD groups (Cowan et al., 2010; Kofler et al., 2010). Taken together, these findings appear to provide insight into meta-analytic evidence that children with ADHD do not perform comparatively worse than non-ADHD children on digit span backward (d = 0.43) than digit span forward (d = 0.47; Martinussen et al., 2005) despite large magnitude deficits on working memory tasks with prominent executive processing demands ($d \geq 2.0$; Kasper et al., 2012). The current findings are also consistent with recent evidence that WISC-IV digit span classifies ADHD and Non-ADHD children at chance levels (Colbert & Bo, 2017; Tarle et al., 2017), whereas criterion working memory tasks that include increased central executive demands, a greater number of trials, and partial-credit scoring methods demonstrate group classification accuracy at levels similar to or better than gold standard ADHD rating scales (Tarle et al., 2017).

Limitations

Generalization of findings from highly controlled laboratory experiments are always limited, and no conclusions regarding working memory *deficits* in ADHD can be made due to the sample's normatively average digit span scaled scores (Table 1) and the lack of a local comparison group. Thus, although our results were highly consistent with findings from diverse child and adult samples (e.g., Engle et al., 1999), and with the test's standardization sample (Wechsler, 2003), replication with larger samples and other clinical groups is clearly needed. Further, our estimate of working memory updating was based on three blocks of a single task, rather than the multiple tasks available to estimate working memory reordering and dual-processing. As such, its relatively lower predictive and construct validity evidence may be due in part to our inability to remove task-specific error (Engle et al., 1999). Performance on digit span backward may have been influenced by our modification of the digit span forward subtest, which increased the total length of the task. Further studies may care to investigate whether omitting low list-length trials or using computerized administration to adaptively adjust list length (Wood et al., 2011) may shorten administration while retaining construct relevance.

Clinical and Research Implications

While digit span backward's list reversal demands are indeed more difficult than rote recall, it appears that this difficulty prevents many children from reaching list-lengths that challenge working memory by exceeding short-term memory demands. Combined with consistent findings from diverse child and adult samples, these findings challenge key components of our most commonly used neuropsychological and intellectual test batteries, and encourage caution when interpreting digit span performance. At the same time, our findings provide initial evidence that modified discontinuation and scoring rules may enable backward span tests to maximally capture meaningful working memory processes, and should be considered in future test development and revisions. Importantly, we found that improved construct and predictive validities require the administration of high list-length trials *and* partial-credit scoring. The unfortunate implication of this finding is that that large, existing datasets cannot simply be rescored to quickly update the knowledge base. Although the alternate administration and scoring rules may be helpful for researchers, the current lack of normative data limits this method's immediate clinical utility. More broadly, to our knowledge there are few neuropsychological test batteries that adequately measure children's working memory according to the task parameter, administration, and scoring principles identified here[2].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by a UVa Curry School of Education Foundation grant (PI: Kofler) from the Galant Family and an NIH grant (R34 MH102499-01, PI: Kofler). The sponsors had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript.

References

Ackerman PL, Beier ME, Boyle MO. Working memory and intelligence: The same or different constructs? Psychological Bulletin. 2005; 131(1):30–60. [PubMed: 15631550]

Alderson R, Kasper L, Patros C, Lea S. Working memory deficits in boys with ADHD: An examination of orthographic and episodic buffer processes. Child Neuropsychology. 2015; 21:509–30. [PubMed: 24830472]

Alderson RM, Rapport MD, Hudec KL, Sarver DE, Kofler MJ. Competing core processes in ADHD: Do working memory deficiencies underlie behavioral inhibition deficits? Journal of Abnormal Child Psychology. 2010; 38:497–507. [PubMed: 20140491]

[2]Redick, Unsworth, Kelly, & Engle (2012) provide normative data from over 6,000 adults for several freely available criterion working memory dual-processing (complex span) tasks. For children, some but not all Automated Working Memory Assessment (Alloway, 2007) "working memory" subtests meet consensus criteria as working memory tests (Conway et al., 2005), but U.S. norms are not available and the test has been discontinued by the publisher. The NIH Cognitive Toolbox (List Sorting), Woodcock-Johnson-IV-Cognitive (Object-Number Sequencing), and Differential Ability Scales-II (Recall of Sequential Order) each contain one face valid working memory subtest; however, all three may be limited by early discontinue rules and we are unaware of any studies assessing their construct or predictive validities.

Alloway TP, Gathercole SE. The role of sentence recall in reading and language skills of children with learning difficulties. Learning & Individual Differences. 2005; 15:271–282.

Alloway TP, Gathercole SE, Pickering SJ. Verbal and visuospatial short-term and working memory in children: Are they separable? Child Development. 2006; 77:1698–1716. [PubMed: 17107455]

Baddeley, A. Working memory, thought, and action. New York: Oxford; 2007.

Bowden SC, Petrauskas VM, Bardenhagen FJ, Meade CE, Simpson LC. Exploring the dimensionality of digit span. Assessment. 2013; 20:188–198. [PubMed: 22914758]

Colbert AM, Bo J. Evaluating relationships among clinical working memory assessment and inattentive and hyperactive/impulsive behaviors in a community sample of children. Research in Developmental Disabilities. 2017; 66:34–43. [PubMed: 28582687]

Colom R, Flores-Mendoza C, Quiroga MÁ, Privado J. Working memory and general intelligence: The role of short-term storage. Personality & Individual Differences. 2005; 39:1005–1014.

Conway AR, Kane MJ, Bunting MF, Engle RW. Working memory span tasks: A methodological review and user's guide. Psychonomic Bulletin & Review. 2005; 12:769–786.

Conklin HM, Curtis CE, Katsanis J, Iacono WG. Verbal working memory impairment in schizophrenia patients and their first-degree relatives: Evidence from the Digit Span task. American Journal of Psychiatry. 2000; 157:275–277. [PubMed: 10671401]

Cowan N. Metatheory of storage capacity limits. Behavioral and brain sciences. 2001; 24:154–176.

Cowan N, Morey CC, AuBuchon AM, Gilchrist AL. Seven-year-olds allocate attention like adults unless working memory is overloaded. Developmental science. 2010; 13:120–133. [PubMed: 20121868]

Dennis M, Francis D, Cirino P, Fletcher J. Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. J International Neuropsychological Society. 2009; 15:331–343.

Denney CB, Rapport MD, Chung KM. Interactions of task and subject variables among continuous performance tests. Journal of Child Psychology and Psychiatry. 2005; 46:420–435. [PubMed: 15819651]

DiStefano C, Zhu M, Mindrila D. Understanding and using factor scores: considerations for the applied researcher. Practical Assessment, Research & Evaluation. 2009; 14(20):1–11.

Egeland J. Measuring working memory with Digit Span and Letter-Number Sequencing subtests from the WAIS-IV: Too low manipulation load and risk for underestimating modality effects. Applied Neuropsychology: Adult. 2015; 22(6):445–451.

Engle RW, Tuholski SW, Laughlin JE, Conway AR. Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. Journal of Experimental Psychology: General. 1999; 128(3):309.

Fassbender C, Schweitzer JB. Is there evidence for neural compensation in ADHD? A review of the functional neuroimaging literature. Clinical Psychology Review. 2006; 26:445–465. [PubMed: 16500007]

Faul F, Erdfelder E, Lang AG, Buchner A. G* Power 3. Behavior Research Methods. 2007; 39(2): 175–191. [PubMed: 17695343]

Friedman NP, Miyake A. Unity and diversity of executive functions: Individual differences as a window on cognitive structure. Cortex. 2016 [PubMed: 27251123]

Gadow, K., Sprafkin, J. Child symptom inventory 4: Screening and norms. Checkmate Plus: 2002.

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. [PubMed: 14979759]

Goldman-Rakic PS. Working memory dysfunction in schizophrenia. The Journal of Neuropsychiatry and Clinical Neurosciences. 1994; 6:348–357. [PubMed: 7841806]

Hogarty KY, Hines CV, Kromrey JD, Ferron JM, Mumford KR. The quality factor solutions in exploratory factor analysis: the influence of sample size, communality, and over-determination. Educational and Psychological Measurement. 2005; 65:202–226.

Hollingshead, AB. Four factor index of social status. New Haven, Conn: Yale; 1975. 1975

Jaroslawska AJ, Gathercole SE, Logie MR, Holmes J. Following instructions in a virtual school: Does working memory play a role? Memory & Cognition. 2016; 44(4):1–10.

Joormann J, Gotlib IH. Updating the contents of working memory in depression: Interference from irrelevant negative material. Journal of Abnormal Psychology. 2008; 117(1):182. [PubMed: 18266496]

Kane M, Brown L, Kwapil T. For whom the mind wanders: An experience-sampling study of working memory & executive control in daily life. Psychological Science. 2007; 18:614–621. [PubMed: 17614870]

Kaufman J, Birmaher B, Brent D, Ryan N. Schedule for affective disorders and schizophrenia for school-age children (K-SADS-PL): Initial reliability and validity data. Journal of the American Academy of Child and Adolescent Psychiatry. 1997; 36:980–988. [PubMed: 9204677]

Kasper LJ, Alderson RM, Hudec KL. Moderators of working memory deficits in children with ADHD: A meta-analytic review. Clinical Psychology Review. 2012; 32:605–617. [PubMed: 22917740]

Kofler MJ, Rapport MD, Bolden J, Sarver DE, Raiker JS. ADHD and working memory: The impact of central executive deficits and exceeding storage/rehearsal capacity on observed inattentive behavior. Journal of Abnormal Child Psychology. 2010; 38:149–161. [PubMed: 19787447]

Kofler MJ, Rapport MD, Bolden J, Alderson RM. Working memory deficits and social problems in children with ADHD. Journal of Abnormal Child Psychology. 2011; 39:805–817. [PubMed: 21468668]

Kuntsi J, Oosterlaan J, Stevenson J. Psychological mechanisms in hyperactivity: I Response inhibition deficit, working memory impairment, delay aversion, or something else? Journal of Child Psychology and Psychiatry. 2001; 42(2):199–210. [PubMed: 11280416]

Lee IA, Preacher KJ. Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. 2013

Martinussen R, Hayden J, Hogg-Johnson S, Tannock R. A meta-analysis of working memory impairments in children with ADHD. J Am Acad Child Adol Psychiatry. 2005; 44:377–384.

McInnes A, Humphries T, Hogg-Johnson S, Tannock R. Listening comprehension and working memory are impaired in attention-deficit hyperactivity disorder irrespective of language impairment. Journal of Abnormal Child Psychology. 2003; 31:427–443. [PubMed: 12831231]

Miyake A, Friedman NP. The nature and organization of individual differences in executive functions. Current Directions in Psychological Science. 2012; 21:8–14. [PubMed: 22773897]

Monaco M, Costa A, Caltagirone C, Carlesimo GA. Forward and backward span for verbal and visuo-spatial data: standardization and normative data from an Italian adult population. Neurological Sciences. 2013; 34(5):749–754. [PubMed: 22689311]

Nee DE, Brown JW, Askren MK, Berman MG, Jonides J. A meta-analysis of executive components of working memory. Cerebral Cotrex. 2013; 23:264–282.

Patros CH, Alderson RM, Lea SE, Tarle SJ. Context influences decision-making in boys with attention-deficit/hyperactivity disorder: A comparison of traditional and novel choice-impulsivity paradigms. Child Neuropsychology. 2015:1–13. (ahead-of-print).

Perry W, Heaton RK, Potterat E, Braff DL. Working memory in schizophrenia: Transient "online" storage versus executive functioning. Schizophrenia Bulletin. 2001; 27:157. [PubMed: 11215544]

Raiker JS, Rapport MD, Kofler MJ, Sarver DE. Objectively-measured impulsivity and ADHD: Testing competing predictions from the working memory and inhibition models of ADHD. Journal of Abnormal Child Psychology. 2012; 40:699–713. [PubMed: 22271141]

Raghubar KP, Barnes MA, Hecht SA. Working memory and mathematics: A review. Learning & individual differences. 2010; 20:110–122.

Rapport MD, Bolden J, Kofler MJ, Sarver DE, Raiker JS, Alderson RM. Hyperactivity in boys with ADHD: A ubiquitous core symptom or manifestation of working memory deficits? Journal of Abnormal Child Psychology. 2009; 37:521–534. [PubMed: 19083090]

Rapport MD, Orban SA, Kofler MJ, Friedman LM. Do programs designed to train working memory, other executive functions, and attention benefit children with ADHD? A meta-analytic review. Clinical psychology review. 2013; 33:1237–1252. [PubMed: 24120258]

Redick TS, Lindsey DR. Complex span and n-back measures of working memory: A meta-analysis. Psychonomic Bulletin & Review. 2013; 20(6):1102–1113.

Redick TS, Broadway JM, Meier ME, Engle RW. Measuring working memory capacity with automated complex span tasks. European J Psychological Assessment. 2012; 28:164–171.

Reynolds CR, Kamphaus RW. BASC-2: Behavior assessment system for children. 2004

Rowland AS, Lesesne CA, Abramowitz AJ. The epidemiology of ADHD: A public health view. Mental Retardation and Developmental Disabilities Research Reviews. 2002; 8(3):162–170. [PubMed: 12216060]

Rubia K, Alegria AA, Cubillo AI, Radua J. Effects of stimulants on brain function in ADHD: a systematic review and meta-analysis. Biological psychiatry. 2014; 76(8):616–628. [PubMed: 24314347]

Sarver DE, Rapport MD, Kofler MJ, Raiker J, Friedman L. Hyperactivity in ADHD: Impairing deficit or compensatory behavior? J Abnormal Child Psychology. 2015; 43:1219–1232.

Schmiedek F, Lövdén M, Lindenberger U. Putting complex span, N-back, and other working memory indicators in psychometric context. Frontiers in Psychology. 2014; 5:1475. [PubMed: 25566149]

Sesma HW, Mahone EM, Levine T, Eason SH, Cutting LE. The contribution of executive skills to reading comprehension. Child Neuropsychology. 2009; 15(3):232–246. [PubMed: 18629674]

Shelton JT, Elliott EM, Matthews RA, Hill BD, Gouvier WD. The relationships of working memory, secondary memory, and general fluid intelligence: Working memory is special. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2010; 36(3):813–820. [PubMed: 20438278]

Snyder HR, Miyake A, Hankin BL. Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. Frontiers in Psychology. 2015; 6:1–24.

St Clair-Thompson HL. Backwards digit recall: A measure of short-term memory or working memory? European Journal of Cognitive Psychology. 2010; 22(2):286–296.

St Clair-Thompson HL, Allen RJ. Are forward and backward recall the same? A dual-task study of digit recall. Memory & Cognition. 2013; 41(4):519–532.

Swanson L, Kim K. Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. Intelligence. 2007; 35(2):151–168.

Tarle SJ, Alderson RM, Patros CH, Lea SE, Hudec KL, Arrington EF. ADHD and phonological working memory: Methodological variability affects clinical and experimental performance metrics. Neuropsychology. 2017 Advance online publication. [PubMed: 28277685]

Tillman C, Eninger L, Forssman L, Bohlin G. The relation between working memory components and ADHD symptoms from a developmental perspective. Developmental Neuropsychology. 2011; 36(2):181–198. [PubMed: 21347920]

Twamley EW, Palmer BW, Jeste DV, Taylor MJ, Heaton RK. Transient and executive function working memory in schizophrenia. Schizophrenia Research. 2006; 87:185–190. [PubMed: 16737800]

Unsworth N, Engle RW. Simple and complex memory spans relation to fluid abilities: Evidence from list-length effects. Journal of Memory and Language. 2006; 54:68–80.

Unsworth N, Engle R. On the division of short-term and working memory: examination of simple & complex span & their relation to higher-order abilities. Psychological Bulletin. 2007; 133:1038–1066. [PubMed: 17967093]

Unsworth N, Spillers GJ, Brewer GA. The Contributions of Primary and Secondary Memory to Working Memory Capacity: An Individual Differences Analysis of Immediate Free Recall. Journal of Experimental Psychology: Learning Memory and Cognition. 2010; 36(1):240–247. DOI: 10.1037/a0017739

van Daal J, Verhoeven L, van Leeuwe J, van Balkom H. Working memory limitations in children with severe language impairment. Journal of Communication Disorders. 2008; 41:85–107. [PubMed: 17482204]

Wager TD, Smith EE. Neuroimaging studies of working memory. Cognitive, Affective, & Behavioral Neuroscience. 2003; 3(4):255–274.

Wechsler, D. WASI-II: Wechsler Abbreviated Scales of Intelligence. Pearson: 2011.

Willcutt EG, Pennington BF, Boada R, Olson RK. A comparison of the cognitive deficits in reading disability and ADHD. Journal of Abnormal Psychology. 2001; 110:157–172. [PubMed: 11261391]

Woods DL, Kishiyama MM, Yund EW, Herron TJ, Reed B. Improving digit span assessment of short-term verbal memory. Journal of clinical and experimental neuropsychology. 2011; 33(1):101–111.

Author Manuscript

Author Manuscript

Author Manuscript

What this Paper Adds

Pediatric ADHD is associated with impairments in working memory, but commonly used working memory tests in clinical practice show insufficient sensitivity/specificity to aid in differential diagnosis or inform treatment planning for these children. There is substantial evidence indicating that this may be due to poor construct validity of the most commonly used tests. The current study focuses on digit span backward, arguably the most commonly used test of working memory in the ADHD literature. In a strictly-diagnosed and well-characterized sample of children with ADHD, we replicated previous findings indicating that digit span backwards fails to adequately capture meaningful working memory processes. To our knowledge, this was the first study to replicate these findings relative to a broad battery of tests that assess all three empirically-identified working memory processes (reordering, dual-processing, updating). The study's unique contribution – beyond its literature review that highlights substantial but oft-ignored problems with interpreting digit span backward – is its demonstration that minor modifications to the test's administration and scoring may significantly improve its construct validity and predictive validity. If replicated, these findings have immediate implications for researchers and clinicians who value the test's ease of administration but desire increased precision for measuring working memory. These modifications should be considered in future revisions of tests and protocols that utilize digit span as a working memory measure.

Table 1

Demographic and test performance variables

	ADHD Inattentive		ADHD Combined/Hyperactive		Overall Sample	
	M	**SD**	**M**	**SD**	**M**	**SD**
N (Boys/Girls)	19 (9/10)		15 (14/1)		34 (23/11)	
Age	10.67	1.53	10.08	1.13	10.41	1.38
SES	49.16	9.34	47.97	11.47	48.63	10.19
PRI	108.89	13.09	110.47	17.10	109.35	13.99
BASC-2 Attention Problems (*T-scores*)						
Parent	66.89	8.09	68.33	6.88	67.58	7.60
Teacher	63.68	8.93	60.53	7.87	61.23	9.03
BASC-2 Hyperactivity (*T-scores*)						
Parent	66.47	14.32	79.13	8.98	71.18	13.17
Teacher	54.58	8.24	68.07	12.90	58.70	12.40
Digit Span Traditional Scoring/Discontinue Rules (*Trials Correct*)						
Forward	9.21	2.57	8.87	1.92	9.06	2.28
Backward	6.68	1.64	7.07	1.71	6.85	1.65
Digit Span Partial-Credit Scoring/All Trials Administered (*Stimuli Correct*)						
Forward Total Digits Correct	47.89	18.28	49.27	11.90	48.57	15.20
Forward Low List-lengths	17.84	0.50	17.93	0.26	17.86	0.42
Forward High List-lengths	21.89	16.39	22.60	11.06	22.43	13.81
Backward Total Digits Correct	37.37	7.02	35.87	10.54	37.24	8.61
Backward Low List-lengths	18.26	2.60	18.80	2.51	18.65	2.51
Backward High List-lengths	14.26	4.95	12.27	6.27	13.78	5.68
Working Memory Reordering (*Stimuli Correct Per Trial*)						
PH 3	2.94	0.10	2.96	0.07	2.95	0.09
PH 4	3.33	0.77	3.36	0.58	3.34	0.68
PH 5	3.18	1.32	3.29	1.11	3.23	1.22
PH 6	2.62	1.47	2.78	1.56	2.69	1.49
VS 3	2.28	0.59	2.20	0.58	2.25	0.58
VS 4	2.81	0.72	2.41	0.97	2.63	0.85
VS 5	2.79	1.01	2.49	1.27	2.66	1.12
VS 6	2.42	1.20	2.12	1.63	2.28	1.39
Working Memory Dual-Processing/Complex Span (*Stimuli Correct Per Trial*)						
CS 3	2.46	0.35	2.33	0.33	2.40	0.36
CS 4	2.71	0.19	2.73	0.23	2.72	0.20
CS 5	2.63	0.39	2.63	0.47	2.63	0.42
CS 6	2.28	0.51	2.27	0.58	2.28	0.54
OS 3	2.58	1.49	2.07	0.94	2.35	1.29
OS 4	2.63	0.76	2.80	1.03	2.71	0.88
OS 5	2.79	1.08	3.10	1.62	2.92	1.33

	ADHD Inattentive		ADHD Combined/Hyperactive		Overall Sample	
	M	**SD**	**M**	**SD**	**M**	**SD**
OS 6	2.61	1.28	3.03	1.66	2.79	1.45
Working Memory Updating (*Total Errors*)						
Block 1 Errors	23.42	16.25	31.47	22.64	26.97	19.44
Block 2 Errors	36.11	22.12	45.93	27.53	40.44	24.76
Block 3 Errors	39.68	21.58	45.80	22.20	42.38	21.74

Note. BASC-2 = Behavior Assessment System for Children (T-scores); CS = Counting Span WM (Stimuli Correct/Trial); OS = Operation Span WM (Stimuli Correct/Trial); PH = Phonological WM (Stimuli Correct/Trial); PRI = WASI-II Perceptual Reasoning Index (Standard Scores); VS = Visuospatial WM (Stimuli Correct/Trial); WM = Working Memory.

Table 2

Construct validity of digit span backward as a function of administration and scoring methods.

	DSB Traditional Administration/Scoring	DSB Modified Administration/Scoring	Test of dependent *r* (z_{rdep})
Working Memory Reordering	.03	.58 **	4.38, p < .0005
Working Memory Dual-Processing	.22	.53 **	2.44, p = .01
Working Memory Updating	−.03	.28 **	2.22, p = .03
Fluid IQ	.49 **	.15	2.60, p = .005

Note: Modified administration/scoring includes administering all trials (ignoring discontinue rule) and counting each digit recalled in the correct serial position at high list-length trials (memory sets 6–8). Tests of dependent *r* compare the traditional vs. modified administration/scoring approaches to determine whether their associations with each criterion indicator are significantly different (e.g., did the modified scoring/administration significantly improve DSB's association with working memory?). DSB = digit span backward; WM = working memory.

** 95% confidence interval excludes 0.0, indicating significant effects.

Table 3

Predictive validity of working memory reordering, working memory dual-processing, working memory updating, and WASI-II PRI_{res}

	KTEA-2 Achievement Composite
DSB Traditional Administration and All-or-nothing Scoring	.26
DSB High List-length Trials with Partial-credit Scoring	.49 **
WM Reordering	.59 **
WM Dual-Processing	.73 **
WM Updating	.43 **
Fluid IQ	.14

Note: DSB = digit span backward; WM = working memory. Test of dependent *r* for DSB traditional vs. modified administration/scoring: z_{rdep} = 1.78, *p* = .07.

** 95% confidence interval excludes 0.0, indicating significant effects.