Published in final edited form as:

Dyslexia. 2017 August ; 23(3): 209–233. doi:10.1002/dys.1557.

Short-term Memory in Childhood Dyslexia: Deficient Serial Order in Multiple Modalities

Nelson Cowan¹, Tiffany P. Hogan², Mary Alt³, Samuel Green⁴, Kathryn L. Cabbage², Shara Brinkley⁵, and Shelley Gray⁵

¹Department of Psychological Sciences, University of Missouri, USA

²Department of Communication Sciences and Disorders, MGH Institute of Health Professions, Cambridge, MA, USA

³Department of Speech, Language, and Hearing Sciences, University of Arizona, USA

⁴T. Denny Sanford School of Social and Family Dynamics, Arizona State University, USA

⁵Department of Speech and Hearing Science, Arizona State University, USA

Abstract

In children with dyslexia, deficits in working memory have not been well-specified. We assessed second-grade children with dyslexia, with and without concomitant specific language impairment, and children with typical development. Immediate serial recall of lists of phonological (nonword), lexical (digit), spatial (location), and visual (shape) items were included. For the latter three modalities we used not only standard span but also running span tasks, in which the list length was unpredictable to limit mnemonic strategies. Nonword repetition tests indicated a phonological memory deficit in children with dyslexia alone compared to those with typical development, but this difference vanished when these groups were matched for nonverbal intelligence and language. Theoretically-important deficits in *serial order memory* in dyslexic children, however, persisted relative to matched typically-developing children. The deficits were in recall of (a) spoken digits in both standard and running span tasks, and (b) spatial locations, in running span only. Children with dyslexia with versus without language impairment, when matched on nonverbal intelligence, had comparable serial order memory, but differed in phonology. Because serial orderings of verbal and spatial elements occur in reading, the careful examination of order memory may allow a deeper understanding of dyslexia and its relation to language impairment.

Dyslexia is a well-known disorder in children and adults, in which an individual has difficulty reading despite apparently normal performance in other related domains. We include the word “apparently” to allow that, underlying dyslexia, there may be general information processing deficits that are most evident for reading, but that affect certain other types of performance as well. In particular, it has been suggested recently that individuals with dyslexia have difficulty remembering the serial order of information (for a review, see Majerus & Cowan, 2016). Ramus and Szenkovits (2008, p. 134) noted that children with

dyslexia in their studies had no difficulty remembering and repeating isolated nonwords (a measure of short-term phonological memory) but were deficient in remembering series of two or three nonwords. It would clearly be detrimental to reading if the order of such elements were not well-remembered, but does the ordering deficit apply to real-word stimuli? What is the extent of the materials that produce deficient processing, and under what conditions?

We investigated the nature of short-term memory deficits in second-grade children with a varied set of materials including spoken nonwords, spoken digits, spatial locations, and irregular shapes. Definitions of short-term or working memory vary (Logie & Cowan, 2016), but here we refer to tasks in which a certain series of materials is presented and, when the series ends, the instruction is that the participant is immediately to repeat the series or list using a well-suited response mode. The short-term retention of the order of letters, phonemes, words, and ideas must be important early on in the process of reading, and therefore our study is intended to determine if short-term retention could be impaired for reasons that either are specific to verbal materials or are more general.

Recent Literature on Serial Order Memory Deficits in Dyslexia

Surprisingly little research has investigated whether individuals with dyslexia have trouble remembering serial order (see Majerus & Cowan, 2016). Several previous studies have examined nonverbal, as well as verbal, serial order short-term memory mechanisms in adults who had developmental dyslexia using a variety of methods, and have found some deficits in serial order memory (in Dutch, Hachmann et al., 2014; in Finnish, Laasonen et al., 2012; in French, Martinez Perez et al., 2013, 2015; in English, Romani, Tsouknida, & Olson, 2015; Wang, Xuan, & Jarrold, 2016). There are very few relevant studies with children, and they have differed in suggesting that there is a serial order memory deficit (in French: Martinez Perez, Majerus, Mahot, & Poncelet, 2012) or that there is not (with a similar population: Staels & van den Broeck, 2014, 2015).

The Current Experimental Approach

We believe that the investigation of the serial order issue must rest on a carefully considered philosophical approach. There are both practical and theoretical issues to address. Practically, one wants to know something about the kinds of information processing deficits displayed by children with dyslexia so that these deficits can be addressed. Theoretically, one wants to know if there is a skills profile specific to dyslexia that can be evaluated by matching participant groups on broad skills to look for residual deficits.

Test groups

To obtain the necessary knowledge, we collected data from a large sample of children with typical development (TD group) and with dyslexia. Given that a large proportion of children with dyslexia have concomitant specific language impairment (Catts, Adlof, Hogan, & Weismer, 2005), we divided children with dyslexia into those with co-morbid language impairment (DYS+SLI group) and those without (DYS group). We screened out children with other disorders, such as attention deficits. Interesting group differences emerged.

Matching

For a deeper understanding of dyslexia, it is important to ask whether there is a profile of skills that makes children with dyslexia distinct from others. Toward this end, we matched the TD and DYS groups on both nonverbal intelligence and oral language to determine what DYS group deficits could not be attributed entirely to differences in these common characteristics. Similarly, we matched the DYS and DYS+SLI groups on nonverbal intelligence to determine how characteristics of language impairment contribute to short-term memory.

We matched groups only once and, when a child who was part of a matched pair did not complete a measure, that child and the match were both omitted from the statistical analysis of that measure. It theoretically might be possible to make different matches on a measure-by-measure basis to maximize the sample size for each measure but we avoided that re-matching method, as it could lead to bias (e.g., it circumvents the children who did not pass training on a particular measure, which happened with frequencies as shown in Table 1).

Psychometrics

Finding the right levels of stimuli to observe a differential deficit is a tricky business (MacDonald, in press). We wished to observe a potential deficit in serial order memory beyond any deficit in item memory. To accomplish that, we used the same stimuli in each group, and scored the results by the number of items correct. This measure reaches a maximum level based on the participant's ability and does not continue to increase with further increases in difficulty of the materials beyond the participant's ability level, such as the length of list to be repeated (Cowan, 2001). For good equivalence of item and order memory measures we examined them in the same tasks, expressing order memory as a proportion of item memory. Specifically, the response to each item to be recalled was scored as correct if that item was reproduced at all (in lax scoring) and as correct only if it was reproduced in the correct serial position (in strict scoring). Then the list totals were used to calculate the proportion of loss due to serial order scoring, as a performance ratio (PR): $PR = (\text{number items recalled, lax} - \text{number items recalled, strict}) / \text{number items recalled, lax}$. Increasing values of this ratio indicate an increasing proportion of those items that were recalled at all occurring at the wrong serial positions; thus, lower values indicate better serial order retention. This method was applied to digit, location, and shape span tasks. We considered our method better than the common practice of using separate tasks for item and order measures, which might have produced an underestimate of serial position effects inasmuch as participants often use serial position or serial order for mnemonic support even for tasks that do not require correctly ordered responses, such as free recall (e.g., Ward, Tan, & Grenfell-Essam, 2010).

Finally, to compare participants on phonological memory, which is often alleged to be a fundamental deficit in dyslexia (but see Ramus & Szenkovits, 2008; Szenkovits, Darma, Darcy, & Ramus, 2016), children also completed two nonword repetition tasks (after Catts, 1986; Dollaghan & Campbell, 1998).

Special diagnostic span task types

This research includes an important methodological distinction that can shed light on the nature of group differences. Specifically, the digit, spatial location, and shape span tasks were presented in both standard and running span varieties. In our standard span, the list length was fixed for two trials and then increased by one item, with that protocol repeated up to a predetermined list length (which was determined based on the performance of pilot participants). Thus, the list length was predictable. Running span was developed by Pollack, Johnson, and Knaff (1959) and applied to children by Cohen and Sandberg (1980). In our running span task the list lengths were much longer than participants' capacity and ended at an unpredictable point on each trial. The task at that point was to recall as many items as possible in forward order from the end of the list, ending with the last list item. The critical difference between those two methods is that there are various mnemonic strategies that can be used by participants in standard span (at least with verbal materials, which is what has been used in most past studies), but not easily in running span (Crowder, 1969; Elosúa & Ruiz, 2008; Hockey, 1973; Palladino & Jarrold, 2008). For example, Hockey (1973) found that when the verbal presentation rate was fast, attempting to rehearse the list as it was presented actually impaired performance rather than helping it, compared to waiting passively for the list to end.

It is possible that children with dyslexia would have a deficit in the use of strategies to retain serial order and therefore would show a deficiency specifically on standard span tasks. Conversely, they could have a deficit in a non-strategic aspect of memory that would show up in running span, whereas it might be possible for them to compensate for this deficit strategically in standard span (Cowan et al., 2005). There is also an ecological validity of running span inasmuch as, in spoken or written language comprehension as in running span, the endpoint is unpredictable and a memory load is imposed. In running span using coherent speech as the stimuli, the last phrase is remembered in a rather list-like manner because the verbatim information has not yet been converted into a stable linguistic structure (Jarvella, 1970).

In sum, our main goal is to investigate whether dyslexia is accompanied by a deficit in serial order memory, both in practical terms as examined by whole-group differences, and in theoretical terms as further examined by groups matched for intellectual skills (and, when appropriate, matched for oral language). That matching process compares children with and without dyslexia (without oral language impairment) and then compares children with dyslexia who do or do not have concomitant, specific oral language impairment. To preview the present results, a serial order memory deficit tended to occur in children with dyslexia, with or without specific language impairment, when linguistic materials were used; and it occurred even with non-linguistic materials such as spatial locations, when running memory span methods were used to limit mnemonic strategies.

Method

Overview

Although the estimated prevalence of dyslexia depends on the inclusionary test criteria used in a particular study (as pointed out, for example, by Siegel, 2006), we focus on rather common criteria of dyslexia, including children who passed several screening tests of intelligence, oral language, and normal articulation and had no other disorders such as attention deficit hyperactivity disorder, but scored at or below the 20th percentile on a standard test of phonemic decoding and sight word reading efficiency (Test of Word Reading Efficiency - Second Edition (TOWRE-2; Torgesen, Wagner, & Rashotte, 2012). Because many children with developmental dyslexia exhibit concomitant specific language impairments (SLI; Catts, Adlof, Hogan, & Weismer, 2005), we focused on children with developmental dyslexia with and without language impairment. Together, these groups represent the large majority of children classified with learning disabilities in school systems (Tomblin et al. 1997). It is theoretically plausible that the origins of dyslexia differ when the foundation of primary language is intact versus deficient. For comparison, we also examined a large sample of same-age children with typical development. Along with the span tasks, we administered a battery of standardized speech, oral language, and nonverbal IQ tests to help characterize group differences.

Participants were recruited from schools in Arizona, Nebraska, and Massachusetts in the U.S. They received a \$15 gift card for their participation. On Day 1 participants completed hearing and vision screenings and two tests to identify children with dyslexia, specific language impairment, or dyslexia with concomitant language impairment: the Phonemic Decoding and Sight Word Efficiency subtests of the TOWRE-2 and the core language composite of the Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003). They also completed the Passage Comprehension subtest of the Woodcock Reading Mastery Test (WRMT-III; Woodcock, 2011). Participants who qualified for inclusion in the study on Day 1 completed several other psychometric tests on Day 2: The Expressive Vocabulary Test- Second Edition (EVT-2; Williams, 2007), a nonword repetition task (Dollaghan & Campbell, 1998), and the nonverbal scale of the Kaufman Assessment Battery for Children - Second Edition (KABC-2; Kaufman & Kaufman, 2004).

The span tasks reported here were part of the *Comprehensive Assessment Battery for Children – Working Memory* (CABC-WM; Gray, Alt, Hogan, Green, & Cowan, under development; Cabbage et al., 2016) that includes working memory and executive function tasks. This battery has previously been used (Gray et al., 2016) to assess the structure of working memory in children with typical development who are further described here. The CABC-WM was programmed in Adobe Flash software and administered on a Toshiba DX1210-ST4N22 desktop computer with a 21.5-inch touchscreen monitor. The CABC-WM was set in the context of a pirate adventure game in which a Pirate character narrated instructions throughout the entire set of tasks to motivate children. Testing on the working memory and executive function tasks was carried out in six separate sessions. From this series of tasks, we selected the six span tasks for the present detailed analyses and

comparisons with reading, oral language, and nonverbal IQ measures. Omitted here are three memory tasks (two n-back and one updating) in which memory and executive processing are confounded, as well as several executive function tasks (Stroop, global/local, and response-conflict tasks) that did not produce a reliable, consistent, interpretable pattern of results.

Participants

The characteristics of the three groups of children are shown in Table 1. These children are drawn from the POWWER study (see author notes), which included the groups reported here as well as Spanish-English bilingual children and children with specific language impairment without dyslexia. Participants completed a total of 6 word learning games and a comprehensive battery of working memory tasks, over the course of at least 6 days.

There were multiple inclusionary criteria for the group of children with typical development (TD group): (a) an age between 7 and 9 years; (b) enrollment in second grade or just completed that grade; (c) normal bilateral hearing at 20 dB for frequencies of 1, 2, and 4 kHz; (d) normal or corrected-to-normal vision with an acuity of at least 20/40 on a Lea symbols card; (e) normal color vision per screening (Waggoner, 2002, Color Vision Testing Made Easy, Gulf Breeze, FL), (f) absence of a history of neuropsychiatric disorders such as ADHD or autism spectrum disorder; (g) absence of any history of special education services; (h) English as the first and only language spoken; (i) absence of having repeated a grade; (j) a minimum standard score at the 31st percentile on the Goldman-Fristoe Test of Articulation – 2 (GFTA-2: Goldman & Fristoe, 2000) except for scores below that percentile resulting from errors on a single consonant sound; (j) a minimum standard score of 75 on the KABC-2; (k) a second-grade minimum composite standard score of 96 on the TOWRE-2; and (l) a minimum standard score of 88 on the CELF-4.

Inclusionary criteria for the other groups were similar but there were several key differences. First, children in these groups were not excluded for having repeated a grade, provided that they were between 7 and 9 years of age. Second, they were not excluded for a history of special education services. Third, children were classified as having dyslexia (the dyslexia-only or DYS group) if they had a TOWRE-2 standard score of 88 or lower (20th percentile) and a CELF-4 standard score of at least 88. We chose 20th percentile cutoff scores for consistency inasmuch as the diagnostic criterion for dyslexia varies widely across studies, ranging from the 7th percentile (Badian et al., 1990) to the 30th percentile (Manis et al., 1996). Fourth, children were classified as having specific language impairment if they had a CELF-4 score of 82 or lower, reflecting one standard deviation below the mean plus one SEM below that. This diagnostic criterion has a sensitivity of 100% and specificity of 82% on this measure (Semel et al., 2003). There were not enough children with specific language impairment alone to include in the study (only 9 of them). However, children who met the criteria for both dyslexia and specific language impairment (the DYS+SLI group) thus had a TOWRE-2 standard score of 88 or lower and a CELF-4 standard score of 82 or lower.

Children who met age and grade requirements but were omitted from the present data set included 61 who were Spanish-English bilinguals with TD and 8 who had language impairment without dyslexia, neither of which were within the focus of the present study.

Omitted participants also included 52 who scored in an intermediate range on the criterion tests (the TOWRE-2, the CELF-4, or both), too low to be included as typically developing (TD) but too high to be included in the other two groups of this study (DYS, DYS+SLI), and others whose scores were below cutoff on the GFTA-2 ($n=4$), the KABC-2 ($n=2$), hearing screening ($n=12$), or color vision screening ($n=15$).

Apparatus, Stimuli, and Procedure

Each child was seated 52 cm in front of a computer touch-screen in a quiet workspace. At that distance, each 11.5-cm arrow in the location span tasks subtended 13 degrees of visual angle, and each item in the shape span tasks fit into a 14-cm $\times$ 10-cm rectangle that subtended 16 degrees in width and 11.5 degrees in height. Children sat next to a trained research assistant and both wore headphones, with the sounds adjusted to a standard, comfortable listening level. The child's headphones were also equipped with microphones.

Nonword repetition span tasks—We administered the nonword repetition task of Dollaghan and Campbell (1998) as noted above, and also a novel nonword repetition task. In our novel task, all nonwords had two to five syllables and each syllable was a consonant-vowel-consonant sequence. For example, the structure of each trisyllabic nonword in terms of consonants (C) and vowels (V) was CVCCVCCVC (e.g., in English transliteration, *yitvadgum*). The task was to listen to each nonword and then repeat it verbatim. A nonword was counted as correctly repeated if and only if all consonants were deemed to be repeated correctly and in the in the correct sequence. If a child substituted one sound for another consistently in the GFTA-2 test, then the same substitution was counted as correct in nonword repetition. There were three practice trials with 2 syllables, followed by four trials each of 2-, 3-, 4, and 5-syllable nonwords. Inter-rater reliability was estimated at 87%. In contrast to our procedure, Dollaghan and Campbell did not use a consistent CVC syllabic structure (CVCCVC...), but rather usually an alternation of consonants and vowels. We report the proportion of consonants correct, in the correct serial positions, for both measures.

List span tasks—For each of the six list span tasks, children completed a training sequence prior to the testing phase to ensure understanding of the task. If a child could not pass the training sequence after five attempts, the task was discontinued. See Table 1 for a summary of how many children were excluded. In each type of span, a presentation rate of 1 s per stimulus was used. The response selection was by touch screen for nonverbal stimuli and was spoken for digit stimuli, with the experimenter keying in the correct response (with audio backup to check reliability). In the response on standard span tasks, the computer automatically advanced to the next trial when the requisite number of responses were provided but, in the response on running span tasks, the participant (or for digit spans, the experimenter) pressed a key to indicate that the response was finished, advancing the computer to the next trial. Certain other aspects of the procedure were necessarily different between tasks (see Fig. 1) based on pilot data that informed the nature of practice and the list lengths needed to capture the children's range of abilities. The span types were administered in random order among other cognitive tests of working memory, executive control, and word learning. For each span task, cumulative feedback was provided at the end of the task

in the form of virtual gold coins for each correct response and virtual rocks for each incorrect response.

Standard digit span—A training sequence included two trials of 2-digit lists. Then the test included 14 trials: two each with List Lengths 2–8, starting with the shortest sequences and steadily increasing in length. Each trial began with the simultaneous appearance of a cartoonish character (Fig. 1), simultaneous with the presentation of the first digit in the list, to encourage the impression that the digits were spoken by this character and thus were still within the pirate game. Digits were presented by computer in a recorded male voice with a monotonic intonation, in a different random order on every trial. The digit 7 was always omitted because it is the only digit with two syllables. At the end of the list, a green rectangle served as a cue for the child to try to recall the list orally; the computer audio-recorded the response for backup, and the research assistant wrote down the response and then entered it into the computer. This test procedure took, on average, 4.5 minutes including instructions and practice.

Running digit span—On every trial seven to ten digits were presented, with the endpoint unknown to the participant. Again, the digit 7 was always omitted. The task was to repeat as many items as possible from the end of the list, in forward order, after the list ended. Successful training required that the participant get 3 trials in a row correct, and training was failed if the participant got 5 in a row incorrect. (Correct responses were defined as at least the last item in the list correct.) After training was passed, there were 12 test trials (3 per list length, with lengths randomly intermixed). The responses were spoken by the child and audio-recorded by the computer for backup as well as keyed in by the research assistant. The procedure took approximately 6 minutes to complete.

Standard location span—As shown in Fig. 1, the task involved the presentation of 11.5-cm-long arrows that originated at the center of the screen and pointed, one at a time, to dots that remained in place around a center circle in eight difficult-to-name locations. The task was to use the touch screen to recreate a sequence of dots to which the arrows had pointed. Training included three trials (a trial at List Length 1 and two trials at List Length 2), of which two trials had to be correct to pass training. The scored task included ten test trials, consisting of two trials at each list length (2–6) in ascending order. The end of each response series initiated the next trial. The task took about 4.5 minutes to complete.

Running location span—This task was similar to the standard location span except that, on every trial, there were 5–8 arrows in sequence, pointing to different locations, and the number of locations was not known by the participant; the list lengths were randomized. The task was to use the touch screen to recall as many locations as the child thought he or she remembered from the end of the list, in forward order. The child pressed an “I’m done” button when he or she completed each list. Training requirements were the same as in running digit span (three consecutive trials with at least the last item correct meant training was passed; five consecutive incorrect trials meant training was failed). There were then 12 scored test trials (three per list length, with order randomized). The test took, on average, about 7.5 minutes to complete.

Standard shape span—The stimuli for this task were series of six irregular polygons that would be difficult to name (Fig. 1). They were all presented in the middle of the screen and fit into an area of 14×10 cm. After a series was presented, the task was to select the presented shapes in the presented order from a field of six shapes shown on the touchscreen. A training sequence included a 1-object trial and a 2-object trial, both of which had to be answered correctly to proceed. The scored test included 12 trials, two each of 1–6 polygons, with list length steadily increasing. The task took about 6.5 minutes to complete.

Running shape span—This task was similar to the standard shape span except that each trial included lists of 3–6 polygons in random order so that the endpoint was unpredictable to the participant. As in the location span task, children pressed an “I’m done” button to indicate when they were finished with a list. Training requirements were as in the other running span tasks. The scored test included 12 test trials (three per list length) taking on average about seven minutes to complete.

Bayesian statistical analysis—In this article we use Bayesian inferential statistics (JASP software: Love et al., 2015; *t* tests: Rouder, Speckman, Sun, Morey, & Iverson, 2009; ANOVA: Rouder, Morey, Speckman, & Province, 2012), which do not have the same shortcomings as the null hypothesis significance testing method. Instead of evaluating the likelihood of the data under the null hypothesis, which is not really the theoretically desired quantity, these methods assess a Bayes Factor (BF) ratio, which is the likelihood of a weighted collection of reasonable alternative hypotheses compared to the null, given the data. Ratios greater than 3 are generally considered non-negligible support for the alternative hypothesis set, and ratios less than 1/3 provide comparable evidence favoring the null. To facilitate understanding of the findings, BF ratios are accompanied by the traditionally-calculated *t* and *F* values and effect sizes to describe the data.

Results

Table 1 details the psychometric characteristics of children in the TD, DYS, and DYS+SLI groups and shows that these groups were matched on key characteristics. We first present results based on the entire sample to describe the data landscape, and then examine memory in subsamples matched for scores on nonverbal intelligence and oral language measures. The purpose of the matching method is to identify specific mechanisms of dyslexia that do not depend on these correlated characteristics.

Whole-sample Characteristics

Fig. 2 shows the relation of nonverbal intelligence (measured by the KABC- 2 score) to the criterion measure for dyslexia (left panel) and to the criterion measure for language impairment (right panel). The purpose of these graphs is to show how the classificatory variables are related to intelligence in each group. We present this information because we believe that a careful evaluation of the effects of specific variables occurs best within the general context of the characterization of children. That characterization begins with Table 1 and is continued with Fig. 2.

Clearly, the groups are not equivalent on nonverbal intelligence, though there is considerable overlap, and the TD versus DYS groups are not quite equivalent on oral language; oral language was by definition much worse in the DYS+SLI group but, on average, this group also had lower nonverbal intelligence in our sample, as would be predicted from other groups of children with specific language impairment (Gallinat & Spaulding, 2014). Memory characteristics of the three groups of children differ and are depicted in the supplementary figures (Supp1-Supp5).

Breakdown of intelligence test scores

The KABC-2 scores are based on several subtests, and therefore contain important information about what skills differentiate the groups of children. An examination of each of these subtests indicates that the TD vs. DYS groups differed on the basis of subtests that were not the same as the ones that best distinguished the DYS vs. DYS+SLI groups. One subtest specifically depended on serial order, the *hand movements* subtest, in which the examinee was to imitate a sequence of fist, palm, and side-of-hand positions on the table. That subtest was the one that best differentiated between children with TD ($M=11.22$, $SD=2.68$) vs. those with dyslexia ($M=9.49$, $SD=2.29$), $t(247)=5.00$, $d=0.69$, $BF=12,540$. That test, however, somewhat poorly discriminated between children with dyslexia alone and those with dyslexia and concomitant SLI ($M=8.50$, $SD=1.98$), $t(124)=2.41$, $d=0.46$, $BF=2.63$. Another subtest that discriminated between TD and DYS groups almost as well was *pattern reasoning*, in which a choice must be selected to complete a pattern (e.g., XOX, XOXO, ?, XOXOXO, XOXOXOX, the correct selection out of several being XOXOX; TD: $M=13.18$, $SD=3.07$; DYS: $M=11.32$, $SD=3.09$; $t(247)=4.49$, $d=0.60$, $BF=1,489$). This variable also discriminated moderately between DYS and DYS+SLI ($M=9.89$, $SD=3.04$), $t(124)=2.49$, $d=0.47$, $BF=3.11$. A third subtest that discriminated between TD and DYS groups was the *triangles* test, in which the triangles had to be put together to reproduce a pattern (TD: $M=12.22$, $SD=2.93$; DYS: $M=10.62$, $SD=2.86$; $t(247)=4.07$, $d=0.55$, $BF=295$). The DYS+SLI group ($M=9.39$, $SD=2.73$) was not much different from the DYS group in this test, $t(124)=2.35$, $d=0.44$, $BF=2.30$. Finally, the two tests that discriminated least between the TD and DYS groups discriminated best between DYS and DYS+SLI groups. In one of these, *story completion*, a story is told in pictures and the child must select the pictures that best complete the story [TD: $M=12.41$, $SD=2.69$; DYS: $M=11.56$, $SD=2.49$, TD vs. DYS $t(246)=2.40$, $d=0.33$, $BF=2.16$. DYS+SLI: $M=10.11$, $SD=2.83$; DYS vs. DYS+SLI, $t(124)=2.96$, $d=0.54$, $BF=9.57$]. In the other, *block counting*, a structure is shown composed of blocks, some of which are partly or completely concealed but still must be counted [TD: $M=12.11$, $SD=2.32$; DYS: $M=11.60$, $SD=2.38$, TD vs. DYS $t(247)=1.64$, $d=0.22$, $BF=0.52$. DYS+SLI: $M=10.11$, $SD=3.19$; DYS vs. DYS+SLI, $t(124)=2.95$, $d=0.53$, $BF=9.32$.] It seems notable that the subtests that distinguish between TD and DYS groups best (hand movements, pattern reasoning), appear to depend on memory more than the ones that distinguish them least, but best distinguish the DYS vs. DYS+SLI groups (story completion, block counting).

Phonological memory—As our purest measures of phonological memory we had two nonword repetition tests. First, with the Dollaghan and Campbell (1998) test, we had 162 children in the TD group and 75 children in the DYS group. The TD scores ($M=.86$, $SD=.$

09) were higher than the DYS scores, ($M=.82$, $SD=.11$), $t(235)=2.93$, $d=0.40$, $BF=8.18$. There were 41 children with DYS+SLI for this measure ($M=.73$, $SD=.11$), and they scored lower than the children with DYS, $t(114)=4.31$, $d=0.83$, $BF=564$. The results were similar for our own nonword repetition task. We had 159 children with TD and 75 children with DYS with scores and found that the TD scores ($M=.73$, $SD=.14$) exceeded the DYS scores ($M=.67$, $SD=.13$), $t(232)=3.02$, $d=0.43$, $BF=10.49$. The 28 participants with DYS+SLI who had this test scored even lower ($M=.52$, $SD=.11$) compared to the DYS group, $t(101)=5.47$, $d=1.25$, $BF=37,038$. We will see below that the differences between TD and DYS group scores on these measures become more comparable when matched for nonverbal intelligence and language, but that other differences between the groups related to serial order memory remained.

To understand whether there is a working-memory component of atypical development that is not dependent on the level of nonverbal intelligence (or of oral language, when examining dyslexia), we endeavored to match the groups. First and foremost, we present comparisons of the TD and DYS groups, and then a brief comparison of the DYS and DYS+SLI groups.

Matched-Sample Findings, TD versus DYS Groups

The children with TD and DYS were matched on both nonverbal intelligence and oral language using a version of the propensity scoring method (West et al., 2014). Two available language scores (EVT-2 and CELF-4) were converted to z scores and the z scores averaged. The KABC-2 nonverbal intelligence scores were also converted to z scores and a 2-dimensional space was constructed with the language composite and intelligence as the dimensions. Nearest neighbors in this space could be identified; a nearest-neighbor TD-DYS participant pair was one for which no other DYS participant was closer to this TD participant and no other TD participant was closer to this DYS participant. (In case of occasional ties, the first-run participants were used.) After all such nearest neighbors were removed from the data set, the identification of nearest neighbors among the remaining participants was carried out again. The results of 3 waves of nearest-neighbor identification made up the matched set. Fig. 3 shows the 2-dimensional space with all matched pairs (left panel) and with participants who remained unmatched after 3 waves. These remaining participants were non-overlapping in the two-dimensional space and were therefore excluded from the matched comparisons. The result was 67 matched pairs.

Bayesian paired t tests for the main psychometric tests indicated that TD-DYS matching succeeded. BF for both oral language tests was much less than 1, suggesting no language differences between the TD and DYS groups [$t(66)=0.80$, $d=0.10$, $BF=.18$ for the EVT-2 and $t(66)=0.97$, $d=0.12$, $BF=.21$ for the CELF-4 or, inverting these ratios, 5.56 and 4.76 to 1 in favor of the null hypothesis, respectively]. The mean language CELF-4 score was 102 ($SD = 8$), and the mean language EVT-2 score was 106 ($SD=10$). Nonverbal intelligence was well-matched; the KABC-2 result was $t(66)=0.93$, $d=0.11$, $BF=.20$, or 5 to 1 in favor of the null. The mean KABC-2 score across groups was 109 ($SD = 13$). In contrast to the matched variables, there were still the expected differences in reading, on both the TOWRE-2 sight word and phonological decoding subscores [$t(66)=20.08$, $d=2.45$, and $t(66)=15.63$, $d=1.91$, respectively, with BF essentially infinite in both cases] and the WRMT-III [$t(66)=5.51$,

$d=0.67$, $BF=23,332$ to 1]. For the two groups the TOWRE-2 reading mean scores were 107 ($SD = 7$) versus 81 ($SD = 6$), and the WRMT-III reading mean scores were 105 ($SD = 10$) versus 95 ($SD = 11$).

Phonological memory—The advantage of children with TD over children with DYS in nonword repetition no longer held with matched samples. There were 60 matched pairs with the Dollaghan and Campbell (1998) scores and the group means were similar (TD, $M=.84$, $SD=.09$; DYS, $M=.82$, $SD=.11$), $t(59)=1.00$, $d=0.13$, $BF=0.23$ or 4.35 to 1 in favor of the null. For our own measure of nonword repetition, which we had for 63 matched pairs, the scores for children with TD ($M=.70$, $SD=.11$) were not notably different from the scores for children with DYS ($M=.67$, $SD=.13$), $t(62)=1.24$, $d=0.16$, $BF=0.29$, in favor of the null by 3.4 to 1. Thus, matching on language and nonverbal intelligence removed the differences between the groups in nonword repetition.

Serial order memory in standard span tasks—To examine serial order memory independent from the role of item memory, in each task we subtracted the strict serial position scoring from the lax scoring, so larger difference scores for a given stimulus condition reflect poorer serial order memory. Fig. 4 illustrates the distribution of responses for each condition. Three of the panels (top left, top right, and bottom left) show that in all three of the standard spans, the distributions were fairly similar in the TD and DYS groups. Nevertheless, we carried out analyses not on the difference scores directly, but on the ratio [lax-strict difference score / lax score] to measure the proportion of memory information that was lost by using strict serial order scoring. For each kind of standard span, a Bayesian ANOVA was carried out with all available pairs with complete data sets. In this analysis a TD-DYS participant pair was counted as a single participant with group as a within-participant factor. In every such analysis, the support for a list length effect (starting at Length 2, at which order errors become possible) was overwhelming, but support for a group difference depended on the modality.

In shape span, with 47 valid pairs, $BF=0.10$, favoring the null effect 10 to 1. This shape span measure included lists 2–5 items long; the 6-item trials were excluded because there were only six choices, so there could be no item errors. The strict scoring of this list length did not show a group effect in a paired t test, $t(45)=1.90$, $d=0.28$, $BF=0.83$ (an indeterminate result). For the list lengths leading to a valid ratio, 2–5, $F(1,46)=0.03$, $\eta_p^2=0.001$, $BF=.08$, or 13 to 1 in favor of the null. For location span, with 65 valid pairs across List Lengths 2–5, $F(1,64)=0.93$, $\eta_p^2=.01$, $BF=0.13$, about 8 to 1 in favor of the null. For digit spans, though, with 60 valid pairs, the group effect $F(1,59)=3.58$, $\eta_p^2=.06$, $BF=3.63$, favoring a group difference. In this analysis, 2-word lists were omitted because of ceiling effects. The difference in ratio that was observed was variable across list lengths (Table 2). Nevertheless, the generally higher ratios in the DYS group indicate slightly greater loss of serial order information for digits than was seen in the TD group. In all three modalities, BF for the group $\times$ list length interaction strongly favored the null as well.

Serial order memory in running span tasks—The bottom right panel of Fig. 4 shows that the group differences were more marked in running span. The horizontal dashed line in this panel assists the comparison of TD children, to the left of the vertical dashed line, and

DYS children, to the right of that line. Ratios were again examined, this time with 2-tailed, paired t tests. For running digit span, with 32 valid pairs, $t(31)=3.31$, $d=0.59$, $BF=15.26$ because the ratio was much lower in the TD group ($M=.13$, $SD=.11$) than in the DYS group ($M=.23$, $SD=.17$). Similarly, in running location span with 52 valid pairs, $t(51)=3.54$, $d=0.49$, $BF=32.27$, because the ratio was much lower in the TD group ($M=.23$, $SD=.21$) than in the DYS group ($M=.38$, $SD=.26$). In shape span, with 34 valid pairs, the trend was indeterminate, $t(33)=2.49$, $d=.43$, $BF=2.63$, even though the ratio was again somewhat lower in the TD group ($M=.31$, $SD=.24$) than in the DYS group ($M=.43$, $SD=.20$).

In sum, at least in standard digit span, running digit span, and running location span, serial order information was lost relative to item information more severely in children with DYS than in children with TD. The differences were most prominent in the running span tasks.

Matched-Sample Findings, DYS versus DYS+SLI Groups

Assessment measures—There were 28 pairs of participants who could be matched for nonverbal IQ (KABC-2) across the DYS and DYS+SLI groups. They were almost perfectly matched on this measure (DYS, $M=98.29$, $SD=13.82$; DYS+SLI, $M=98.29$, $SD=13.52$). They did not differ much in word reading ability according to the TOWRE-2 upon which they were both classified as having dyslexia (DYS, $M=80.54$, $SD=7.50$; DYS+SLI, $M=79.32$, $SD=6.54$), though some broad reading-mastery differences remained according to the WRMT-III (DYS, $M=93.25$, $SD=11.55$; DYS+SLI, $M=83.00$, $SD=11.47$). Naturally, they differed considerably on the CELF-4 measure upon which they were differently classified (DYS, $M=97.82$, $SD=7.73$; DYS+SLI, $M=73.21$, $SD=7.99$), and on the other language measure, the EVT-2 (DYS, $M=99.18$, $SD=11.10$; DYS+SLI, $M=89.79$, $SD=9.37$).

The groups still differed on our nonword repetition measure among 23 pairs with the scores (DYS, $M=.66$, $SD=.12$; DYS+SLI, $M=.52$, $SD=.10$), $t(22)=3.69$, $d=0.77$, $BF=28.75$. Interestingly, on the Dollaghan and Campbell (1998) measure of nonword repetition, which used nonwords without consonant clusters, among 23 pairs with the measure, there was no clear difference between the groups (DYS, $M=.79$, $SD=.13$; DYS+SLI, $M=.73$, $SD=.12$), $t(22)=1.73$, $d=0.39$, $BF=0.79$, indeterminate.

In sum, the children with DYS+SLI had lower abilities than those without SLI even when matched for nonverbal IQ, specifically in some aspects of reading, in language, and in complex aspects of phonology (found in our nonword repetition measure compared to the measure of Dollaghan & Campbell, 1998). The latter could be expected on the basis of prior research on language impairment (e.g., Gathercole & Baddeley, 1990).

Serial order memory—Many individuals within the 28 matched pairs of DYS versus DYS+SLI children failed training on the running digit span task. Table 1 shows that complete data did not exist for over half of the children with DYS+SLI and for many children with DYS as well. As a result, only six of the pre-matched pairs with complete data remained to calculate ratios indicating the loss of information with serial order scoring. For those 6 pairs, the ratios were similar (DYS, $M=.26$, $SD=.23$; DYS+SLI, $M=.30$, $SD=.27$), $t(5)=0.33$, $d=0.13$, $BF=.39$. There were 17 matched pairs with running location span data, and their ratios were again quite similar (DYS, $M=.33$, $SD=.26$; DYS+SLI, $M=.37$, $SD=.34$),

$t(16)=0.33$, $d=0.08$, $BF=0.26$, or 3.85 to 1 in favor of the null. Finally, there were 10 matched pairs with running shape span data, and their ratios were again similar (DYS, $M=.44$, $SD=.26$; DYS+SLI, $M=.53$, $SD=.28$), $t(9)=0.84$, $d=0.26$, $BF=0.41$. Given the low numbers of pairs with complete data, however, we also conducted Bayesian ANCOVAs on the entire, unmatched groups with the KABC-2 measure as a covariate and still found no evidence of a group effect [running digit span, n 's=55 & 14, $F(1,66)=2.08$, $\eta_p^2=0.03$, $BF=0.85$, an indeterminate result; running location span, n 's=68 & 23, $F(1,88)=0.01$, $\eta_p^2=0.00$, $BF=0.25$, or 4 to 1 in favor of the null; running shape span, n 's=62 & 19, $F(1,78)=1.82$, $\eta_p^2=0.02$, $BF=0.57$, indeterminate]. For the standard spans, the data for matched pairs were more nearly complete and Group $\times$ List Length ANOVAs on them provided extremely strong evidence for an effect of list length but against any effect of group or its interaction with list length [group effects: digit span, 25 pairs, $F(1,24)=0.08$, $\eta_p^2=0.00$, $BF=0.09$; location span, 24 pairs, $F(1,23)=0.05$, $d=0.00$, $BF=0.10$; shape span, 16 pairs, $F(1,15)=0.02$, $\eta_p^2=0.00$, $BF=0.14$]. Thus, after taking into account nonverbal intelligence, the DYS and DYS+SLI groups appeared to have comparable memory for serial order, which we already found to be below that of children in the TD group.

Correlations and Regressions

Table 3 shows correlations across all participants between the main psychometric variables and a measure of the mean ratio indicating use of serial order information for each of six working-memory measures. The correlations show that serial order information is moderately related to psychometric test scores.

Although it is clear from t -tests between groups and from the correlations that there are sizeable effects of serial order information, it would be helpful to have a better assessment of the importance of serial order memory as a proportion of between-group variance. In order to get a clearer estimate for how much serial order information relates to dyslexia, we compared it to the contribution of nonverbal intelligence. Specifically, we carried out several discriminant analyses (using Statistica 13 software) to classify children for the presence versus absence of dyslexia (TD vs. DYS or DYS+SLI groups). First, we entered into the model all of the subtests of the KABC-2 except for the hand movement test. The four remaining subtests together accounted for 15% of the classification variance (1-Wilks' Lambda), correctly classifying 83% of TD children and 46% of children with dyslexia with or without SLI. When the six variables in the form of ratios representing serial order memory were added (for each participant, the mean ratio across list lengths for each standard span task, and the ratio for each running span task), the model increased from 15% to 33% of the classification variance accounted for, and increased in correct classification from 83% to 89% of TD children and from 46% to 58% of the children with dyslexia. Even with the KABC-2 variables removed, the ratio scores alone about as well as the KABC-2 scores alone, accounting for 24% of the classification variance and correctly classifying 86% of TD children and 43% of children having dyslexia with or without SLI. Given that cognitive skills are considered to be substantially related to dyslexia, we believe that the relation of serial order information to dyslexia is also substantial.

Discussion

Past work has shown that working memory is deficient in children with dyslexia compared to their typically-developing peers (e.g., Ackerman, Dykman, & Gardner, 1990; Cowan & Alloway, 2009; Spafford, 1989). What still has been uncertain is the details of the working memory deficits. Here we point to substantial group differences in short-term serial order memory, findings that can be of use in several ways. First, there may be a mechanism common to serial order memory and the correct serial order processing of units in reading (perhaps including letters, phonemes, words, and ideas), a correlation that does not imply one particular direction of causation. If so, understanding this common serial ordering mechanism may be important for understanding dyslexia. Second, it is possible that dyslexia occurs in part because children with dyslexia lose track of the order of units they are reading. That is, there could be a causal connection between short-term serial order memory and reading ability. Letter series must be held in working memory until words can be identified, and words and phrases must be held in working memory until the underlying ideas can be identified (Jarvella, 1970). Either poor working memory, poor linguistic processing, or both together (as perhaps in the DYS+SLI group) could disrupt the process of reading to varying degrees.

Large deficits were obtained in the unmatched samples of children. In practical terms, there is no doubt that short-term memory deficits with dyslexia include phonological memory deficits (described in the Results section) as well as short-term memory deficits. Purely phonological deficits as reflected by nonword repetition, however, were not found when the TD and DYS groups were matched for nonverbal intelligence and language, implying that these deficits are more a function of serial order memory issues than solely phonological memory deficits.

What remained after matching were deficits in the DYS and DYS+SLI groups compared to the TD group in the serial order component of memory. There were, moreover, potentially heterogeneous bases of these serial order memory deficits. With digit lists, serial ordering of responses was impaired in the DYS group with both standard and running span tasks, though it was much more pronounced with running span. Also in running span, serial order memory for spatial locations was impaired (Fig. 4). Making these serial order memory deficits more noteworthy, they occurred even after matching the TD and DYS groups on tests that included some serial order memory requirements, such as the sentence recall portion of the CELF-4. The trend looks similar for running shape span (for details of running span results see Fig. 5), but with perhaps too much variability among the DYS group for the trend to be statistically compelling.

Overall, there seem to be factors that ease versus complicate the retention of serial order. The figures within the Supporting Information file show that the use of verbal materials eases this kind of retention; in every group, retention was better with digits than with locations or shapes. The reason why children with dyslexia were slightly disadvantaged relative to children with typical development in serial order within standard digit span might be that children with dyslexia are poorer at the covert verbal rehearsal that has been suggested as a mnemonic strategy that adults use (Camos, Mora, & Oberauer, 2011) and is

just developing at this age in children with typical development (e.g., Cowan et al., 2005; Flavell, Beach, & Chinsky, 1966; Ornstein & Naus, 1978). With nonverbal types of span, there may be no equivalent strategy.

Another type of process appears to be needed to explain why children with dyslexia are poorer than children with typical development at the recall of serial order information within running span tasks. Here it has been suggested that a core type of capacity is being measured because mnemonic strategies such as grouping and rehearsal are not possible (Bunting, Cowan, & Saults, 2006; Cowan, 2001; Cowan et al., 2005; Crowder, 1969; Hockey, 1973). One interpretation of the results is that children with DYS have a lower working memory capacity than children with TD, even when the groups are matched for nonverbal intelligence and language (i.e., that children with dyslexia have immature working memory systems; cf. Cowan, 2016). Given that running span tasks are intensive in the requirement of attention (Cowan et al., 2005; Bunting, Cowan, & Colflesh, 2008), it is possible that these serial order deficits will be linked to subtle attention problems in dyslexia (e.g., Facioetti, Ruffino, Peru, Paganoni, & Chelazzi, 2008; Lobier, Zoubrinetzky, & Valdois, 2012; Smith-Spark & Fisk, 2007; Xia, Hoeft, Zhang, & Shu, 2016), even though our samples were screened to eliminate diagnosed attention deficit disorders and matched in nonverbal intelligence.

Overall, serial order was a problem for the DYS-group children more for digit and location spans than for shape span. It is possible that there is a more general temporal-spatial ordering process involved. In spoken materials the phonological storage processes in typical children are well-suited to extract temporal order, whereas in location span, a temporal sequence can be mentally converted into a spatial path (Cowan, Saults, & Morey, 2006). Series of shapes are not particularly amenable to either temporal or spatial sequential coding. Perhaps for this reason, compared to digits or locations, serial ordering for shapes was particularly poor in all groups (Supporting Information, Figures Supp3–Supp5), a near-floor effect reducing the opportunity for group differences to show up statistically, though apparent group differences in serial ordering of shapes still can be observed in Fig. Supp5.

The findings regarding dyslexia seem in keeping with limited previous work supporting the notion that serial order memory is impaired compared to typically-developing individuals (Hachmann et al., 2014; Laasonen et al., 2012; Martinez Perez et al., 2013, 2015; Romani, Tsouknida, & Olson, 2015; Wang, Xuan, & Jarrold, 2016; for review see Majerus & Cowan, 2016). Of the few relevant studies in children, the findings support the conclusion that there is a serial order memory deficit (Martinez Perez, Majerus, Mahot, & Poncelet, 2012) and refute the suggestion that there is not (Staels & van den Broeck, 2014, 2015). At the same time, given that the magnitude of the serial order memory deficit was dependent on a number of test factors in our study (especially the modality and the predictability of the list length), it is not terribly surprising that, under some circumstances, the deficit has not been observed. We found the strongest effects when verbal materials were combined with a running span test, in which the endpoint of the list was not known to the participant so that any planning ahead was minimized, and to our knowledge no previous study has used running spans to compare DYS, DYS+SLI, and/or TD groups. Moreover, compared to past studies, our effects were accentuated by a measurement procedure in which order memory

performance was viewed as a ratio with item information, so that each child's item memory served as an individual control to highlight order information.

Previous work has shown that children with specific language impairment are poorer than matched children with typical development at retaining serial order information from verbal lists (Gillam, Cowan, & Day, 1995). For this reason, it might have been expected that our DYS+SLI group would perform worse in serial order memory than the DYS group.

Although that is the case in the total group data, matching these two groups on nonverbal intelligence was enough to eliminate the group difference. It is possible that children with higher intelligence figure out strategies to circumvent phonological deficits and therefore do not show up as language impaired, whereas those with lower intelligence are more likely to show phonological impairment (Alt, 2011) and, among those with dyslexia, more likely to be in the DYS+SLI group, which shares the serial ordering deficit with the DYS group.

The corpus of research conducted on serial ordering deficits in short-term memory within children with dyslexia is surprisingly small. The present study helps to illustrate that an analytic view of tasks and dependent measures in order to search for fundamental developmental processes (Cowan, Elliott, & Saults, 2002) can be of use also in distinguishing between clinical groups.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by funding from the National Institutes of Health NIDCD Grant #R01 DC010784 known as the POWWER study; also NICHD R01 HD21338 to Cowan.

We are deeply grateful to the staff, research associates, school administrators, teachers, children, and families who participated. Key personnel included (in alphabetical order) Gary Carstensen, Cecilia Figueroa, Karen Guilmette, Trudy Kuo, Bjorg LeSueur, Annelise Pesch, and Jean Zimmer. Many students also contributed to this work including (in alphabetical order) Genesis Arizmendi, Lauren Baron, Alexander Brown, Nora Schlesinger, Nisha Talanki, and Hui-Chun Yang.

References

- Ackerman PT, Dykman RA, Gardner MY. Counting rate, naming rate, phonological sensitivity, and memory span: Major factors in dyslexia. *Journal of Learning Disabilities*. 1990; 23:325–327. [PubMed: 2341803]
- Alt M. Phonological working memory impairments in children with specific language impairment: Where does the problem lie? *Journal of Communication Disorders*. 2011; 44:173–185. [PubMed: 20943232]
- Badian N, McAnulty G, Duffy G, Als HS. Predication of dyslexia in kindergarten boys. *Annals of Dyslexia*. 1990; 41:221–243.
- Bunting MF, Cowan N, Saults JS. How does running memory span work? *Quarterly Journal of Experimental Psychology*. 2006; 59:1691–1700.
- Bunting MF, Cowan N, Colflesh GH. The deployment of attention in short-term memory tasks: Tradeoffs between immediate and delayed deployment. *Memory & Cognition*. 2008; 36:799–812. [PubMed: 18604962]
- Cabbage KL, Brinkley S, Gray S, Alt M, Cowan N, Green S, Kuo T, Hogan TP. Assessing working memory in children: The Comprehensive Assessment Battery for Children – Working Memory

- (CABC-WM). Journal of Visualized Experiments. 2016 Accepted for publication. revisions pending.
- Camos V, Mora G, Oberauer K. Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Memory & Cognition*. 2011; 39:231–244. [PubMed: 21264630]
- Catts HW. Speech production/phonological deficits in reading-disordered children. *Journal of Learning Disabilities*. 1986; 19:504–508. [PubMed: 3760705]
- Catts HW, Adlof SM, Hogan TP, Weismer S. Dyslexia and specific language impairment: Same or different developmental disorder? *Journal of Speech, Language, and Hearing Research*. 2005; 48:1378–1396.
- Cohen RL, Sandberg T. Intelligence and short term memory: A clandestine relationship. *Intelligence*. 1980; 4:319–331.
- Cowan N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*. 2001; 24:87–185. [PubMed: 11515286]
- Cowan N. Working memory maturation: Can we get at the essence of cognitive growth? *Perspectives on Psychological Science*. 2016; 11:239–264. [PubMed: 26993277]
- Cowan, N., Alloway, T. Development of Working Memory In Childhood. In: Courage, ML., Cowan, N., editors. *The development of memory in infancy and childhood*. Hove, East Sussex, UK: Psychology Press; 2009. p. 303–342.
- Cowan, N., Elliott, EM., Sauls, JS. The search for what is fundamental in the development of working memory. In: Kail, R., Reese, H., editors. *Advances in Child Development and Behavior*. Vol. 29. 2002. p. 1–49.
- Cowan N, Elliott EM, Sauls JS, Morey CC, Mattox S, Hismjatullina A, Conway ARA. On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*. 2005; 51:42–100. [PubMed: 16039935]
- Cowan N, Sauls JS, Morey CC. Development of working memory for verbal-spatial associations. *Journal of Memory and Language*. 2006; 55:274–289. [PubMed: 17387377]
- Crowder RG. Behavioral strategies in immediate memory. *Journal of Verbal Learning and Verbal Behavior*. 1969; 8:524–528.
- Dollaghan C, Campbell TF. Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*. 1998; 41:1136–1146.
- Elosúa MR, Ruiz RM. Absence of hardly pursued updating in a running memory task. *Psychological Research*. 2008; 123:451–460.
- Facoetti A, Ruffino M, Peru A, Paganoni P, Chelazzi L. Sluggish engagement and disengagement of non-spatial attention in dyslexic children. *Cortex*. 2008; 44:1221–1233. [PubMed: 18761136]
- Flavell JH, Beach DH, Chinsky JM. Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development*. 1966; 37:283–299. [PubMed: 5941895]
- Gallinat E, Spaulding TJ. Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. *Journal of Speech, Language, and Hearing Research*. 2014; 57:1363–1382.
- Gathercole SE, Baddeley AD. Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*. 1990; 29:336–360.
- Gillam RB, Cowan N, Day LS. Sequential memory in children with and without language impairment. *Journal of Speech & Hearing Research*. 1995; 38:393–402. [PubMed: 7596105]
- Goldman, R., Fristoe, M. *Goldman-Fristoe Test of Articulation*. Second. Pearson; 2000.
- Gray S, Green S, Alt M, Hogan T, Kuo T, Brinkley S, Cowan N. The structure of working memory in young school-age children and its relation to intelligence. *Journal of Memory and Language*. 2017; 92:183–201. [PubMed: 27990060]
- Hachmann WM, Bogaerts L, Szmalec A, Woumans E, Duyck W, Job R. Short-term memory for order but not for item information is impaired in developmental dyslexia. *Annals of Dyslexia*. 2014; 64:121–136. [PubMed: 24488229]
- Hockey R. Rate of presentation in running memory and direct manipulation of input-processing strategies. *Quarterly Journal of Experimental Psychology*. 1973; 25:104–111.

- Jarvella RJ. Effects of syntax on running memory span for connected discourse. *Psychonomic Science*. 1970; 19:235–236.
- Kaufman, AS., Kaufman, NL. Kaufman Assessment Battery for Children. Second. Circle Pines, MN: American Guidance Service; 2004.
- Laasonen M, Virsu V, Oinonen S, Sandbacka M, Salakari A, Service E. Phonological and sensory short-term memory are correlates and both affected in developmental dyslexia. *Reading and Writing*. 2012; 25:2247–2273.
- Lobier M, Zoubrinetzky R, Valdois S. The visual attention span deficit in dyslexia is visual and not verbal. *Cortex*. 2012; 48:768–773. [PubMed: 21982580]
- Logie RH, Cowan N. Perspectives on working memory: Introduction to the special issue. *Memory & Cognition*. 2015; 43:315–324. [PubMed: 25802028]
- Love J, Selker R, Marsman M, Jamil T, Dropmann D, Verhagen AJ, Ly A, Gronau QF, Smira M, Epskamp S, Matzke D, Wild A, Knight P, Rouder JN, Morey RD, Wagenmakers EJ. JASP (Version 0.7.5)[Computer software]. 2015
- MacDonald, AW, III. Differential deficit. In: Cautin, RL., Lilienfeld, SO., editors. *Encyclopedia of clinical psychology set*. Wiley; (in press)
- Majerus S, Cowan N. Is serial order short-term memory impaired in dyslexia? *Frontiers in Psychology*. 2016; 7:1–8. article 1522. DOI: 10.3389/fpsyg.2016.01522 [PubMed: 26858668]
- Manis FR, McBride-Chang C, Seidenberg MS, Keating P, Doi LM, Munson B, Petersen A. Are speech perception deficits associated with developmental dyslexia? *Journal of Experimental Child Psychology*. 1997; 66:211–235. [PubMed: 9245476]
- Martinez Perez T, Majerus S, Mahot A, Poncelet M. Evidence for a specific impairment of serial order short-term memory in dyslexic children. *Dyslexia*. 2012; 18:94–109. [PubMed: 22389071]
- Martinez Perez T, Majerus S, Poncelet M. Impaired short-term memory for order in adults with dyslexia. *Research in Developmental Disabilities*. 2013; 34:2211–2233. [PubMed: 23644228]
- Martinez Perez T, Poncelet M, Salmon E, Majerus S. Functional alterations in order short-term memory networks in adults with dyslexia. *Developmental Neuropsychology*. 2015; 40:407–429. DOI: 10.1080/87565641.2016.1153098 [PubMed: 27043828]
- Ornstein, PA., Naus, MJ. Rehearsal processes in children's memory. In: Ornstein, PA., editor. *Memory development in children*. Hillsdale, NJ: Erlbaum; 1978. p. 69-99.
- Palladino P, Jarrold C. Do updating tasks involve updating? Evidence from comparisons with immediate serial recall. *Quarterly Journal of Experimental Psychology*. 2008; 61:392–399.
- Pollack I, Johnson IB, Knaff PR. Running memory span. *Journal of Experimental Psychology*. 1959; 57:137–146. [PubMed: 13641585]
- Ramus F, Szenkovits G. What phonological deficit? *The Quarterly Journal of Experimental Psychology*. 2008; 61:129–141. [PubMed: 18038344]
- Romani C, Tsouknida E, Olson A. Encoding order and developmental dyslexia: a family of skills predicting different orthographic components. *Quarterly Journal of Experimental Psychology*. 2015; 68:99–128. [PubMed: 25246235]
- Rouder JN, Morey RD, Speckman PL, Province JM. Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*. 2012; 56:356–374.
- Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. Bayesian t Tests for Accepting and Rejecting the Null Hypothesis. *Psychonomic Bulletin & Review*. 2009; 16:225–237. [PubMed: 19293088]
- Semel, E., Wiig, EH., Secord, WA. *Clinical evaluation of language fundamentals*. fourth. Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company; 2003. (CELF-4)
- Siegel LS. Perspectives on dyslexia. *Paediatrics & Child Health*. 2006; 11:581–587.
- Smith-Spark JH, Fisk JE. Working memory functioning in developmental dyslexia. *Memory*. 2007; 15:34–56. [PubMed: 17479923]
- Spafford CS. Wechsler Digit Span subtest: Diagnostic usefulness with dyslexic children. *Perceptual and Motor Skills*. 1989; 69:115–125. [PubMed: 2780171]
- Staels E, Van den Broeck W. Order short-term memory is not impaired in dyslexia and does not affect orthographic learning. *Frontiers in Human Neurosciences*. 2014; 8:732.

- Staels E, Van den Broeck W. No solid empirical evidence for the SOLID (serial order learning impairment) hypothesis of dyslexia. *Journal of Experimental Psychology: Learning Memory and Cognition*. 2015; 41:650–669.
- Szenkovits G, Darma Q, Darcy I, Ramus F. Exploring dyslexics' phonological deficit II: Phonological grammar. *First Language*. 2016; :1–22. DOI: 10.1177/0142723716648841
- Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*. 1997; 40:1245–1260.
- Torgesen, JK., Wagner, RK., Rashotte, CA. *Test of Word Reading Efficiency*. 2nd. Austin, TX: Pro-ED, Inc; 2012.
- Waggoner, TL. *Color Vision Testing Made Easy*. Gulf Breeze, FL: 2002.
- Wang X, Xuan Y, Jarrold C. Using a process dissociation approach to assess verbal short-term memory for item and order information in a sample of individuals with a self-reported diagnosis of dyslexia. *Frontiers in Psychology*. 2016; 7:208. [PubMed: 26941679]
- Ward G, Tan L, Grenfell-Essam R. Examining the relationship between free recall and immediate serial recall: The effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2010; 36:1207–1241.
- West SG, Cham H, Thoemmes F, Renneberg B, Schulze J, Weiler M. Propensity scores as a basis for equating groups: Basic principles and application in clinical treatment outcome research. *Journal of Consulting and Clinical Psychology*. 2014; 82:906–919. [PubMed: 24708350]
- Williams, K. *Expressive vocabulary test*. Second. Circle Pines, MN: AGS; 2007.
- Woodcock, RW. *Woodcock reading mastery test*. 3rd. San Antonio: Pearson; 2011.
- Xia Z, Hoeft F, Zhang L, Shu H. Neuroanatomical anomalies of dyslexia: Disambiguating the effects of disorder, performance, and maturation. *Neuropsychologia*. 2016; 81:68–78. [PubMed: 26679527]

Figure 1.

Illustration of trial details for three kinds of stimuli. Each sequence progresses from upper left to lower right. Lowest sequence: digit span. (Loudspeakers represent spoken digits presented while the little man appears on the screen.) Middle sequence: location span. Upper sequence: shape span. Each sequence depicts a trial with List Length 2, though the procedures as shown are otherwise descriptive of either standard or running span.

Figure 2.

Scatterplots of each group (graph parameter) showing nonverbal intelligence (x axis) as a function of the classification criterion task (y axis). In both panels, triangles represent TD children, discs represent DYS children, and plus signs represent DYS+SLI children. **Left-hand panel**, dyslexia classification according to the TOWRE-2; **right-hand panel**, additional classification of these children by language impairment according to the CELF-4. Some children fell in between on these criteria and were omitted from further testing.

Figure 3. Process of matching typically-developing children (TD group) with those with dyslexia only (DYS group), on the basis of both nonverbal intelligence (KABC-2) and a composite language score. **Left-hand panel:** children included in the match. **Right-hand panel:** children excluded by the matching process.

Figure 4.

Distributions of (lax - strict serial position scoring) difference scores. **Top left:** standard digit span by list length for typically developing children (TD group, left side of dashed vertical line) and children with dyslexia only (DYS group, right side of dashed vertical line). **Top right:** same thing for standard location span. **Bottom left:** same thing for standard shape span. **Bottom right:** running digit, location, and shape spans for typically developing children (left of the dashed vertical line) and those with dyslexia only (right of the dashed vertical line). Bars demarcate the 25th to 75th quartiles. The small point within each rectangles is the median. Whiskers demarcate plus or minus an amount equal to the interquartile range (except that they cannot be negative for these data), circles are outliers beyond the whiskers, and asterisks are extreme values beyond the whiskers by at least twice the interquartile range. The most prominent group differences were in running spans (lower right panel) though differences also emerged in standard digit span (upper left panel).

Figure 5.

Mean items correct with +95% CL in each kind of running span (X axis) using lax scoring (dark bars) and strict serial position scoring (light bars). Left-hand panel: typically developing (TD group) children. Right-hand panel: those with dyslexia (DYS group). Notice poorer performance with strict scoring in the DYS group.

Table 1

Demographic Characteristics and Test Performance of Each Participant Group

Measure	Group									
	Typically Developing (TD)			Dyslexia Only (DYS)			DYS+SLI			Valid N
	Valid N	Mean	SD	Valid N	Mean	SD	Mean	SD	Mean	
Age in months	167	92.82	4.97	80	94.09	6.10	31	94.19	6.98	31
females	95			46			11			11
Mother's Years of Education	166	15.38	1.66	78	14.79	1.91	29	14.00	1.98	29
Self-reported Hispanic	20			16			10			10
Self-reported non-White	30			17			8			8
KABC-2 (IQ)	167	117.60	15.53	80	106.70	13.54	31	96.35	14.16	31
EVT-2	167	112.39	10.95	80	103.24	10.94	31	88.16	10.99	31
WRMT	167	108.23	9.85	80	93.91	10.77	31	82.42	11.51	31
TOWEE-2 (word reading)	167	109.42	8.45	80	80.49	6.25	31	79.42	7.10	31
CELF-4 (language)	167	108.75	9.59	80	100.14	8.51	31	72.65	8.96	31
Maximum Digit Span (standard)	163	4.66	0.86	74	4.24	0.72	31	3.69	0.74	31
failed or ended	3			4			0			0
Digit Span (running)	115	2.68	0.63	55	2.14	0.68	14	1.50	0.66	14
failed or ended	52			24			17			17
Maximum Location Span (standard)	163	2.97	0.82	77	2.75	0.78	30	2.63	0.84	30
failed or ended	3			3			0			0
Location Span (running)	153	1.50	0.61	68	1.20	0.53	23	1.03	0.58	23
failed or ended	13			10			8			8
Maximum Shape Span (standard)	145	2.50	0.89	67	2.14	0.72	23	1.72	0.50	23
failed or ended	19			11			8			8
Shape Span (running)	131	1.26	0.64	62	0.98	0.54	19	0.80	0.40	19
failed or ended	31			14			12			12

Note. Self-report refers to the report about the child coming from the parent or caretaker bringing the child to the test. All test scores are standardized. Abbreviations: KABC-2=Kaufman Assessment Battery for Children-Second Edition; EVT-2= Expressive Vocabulary Test-2; WRMT-III=Woodcock Reading Mastery Test-Passage Comp Scaled Score, Version III; TOWRE-2=Test of Word Reading Efficiency-2 by grade; CELF-4 = Comprehensive Evaluation of Language Fundamentals-4. Maximum span refers to the mean number of items recalled in correct position, for the list length yielding the best result for each participant. "Failed or ended" refers to absence of data for reasons due to the participant, including training not passed or participant opt-out, but excluding cases of experimenter error.

Table 2

Standard Digit Span Serial Order Memory by List Length for Matched Participant Pairs from the Groups with Typical Development (TD) and Dyslexia (DYS).

List Length	Typical Development	Dyslexia Only	Difference
	M (SD)	M(SD)	
3	.00 (.02)	.01 (.07)	-.01
4	.06 (.13)	.07 (.16)	-.01
5	.19 (.24)	.29 (.25)	-.10
6	.45 (.24)	.53 (.27)	-.08
7	.66 (.20)	.65 (.26)	+.01
8	.71 (.20)	.78 (.21)	-.07

Note. Serial order memory is expressed as the ratio [lax-strict difference score / lax score], with large ratios indicating that there is a greater loss of performance when strict serial position information is required by the scoring method. Two-word lists were excluded because of ceiling effects.

Table 3

Correlations between variables across all groups

	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12
1. TOWRE-2	97.75	16.28												
2. WRMT-III	101.23	13.79	0.64											
3. CELF-4	102.25	14.47	0.54	*										
4. EVT-2	107.05	13.43	0.50	*	0.74	*								
5. Nonword (new)	0.65	0.21	0.25	*	0.33	*	0.37	*						
6. Nonword (Dollegahan)	0.79	0.21	0.18	*	0.22	*	0.21	*	0.22	*				
7. KABC-2	112.10	16.52	0.39	*	0.46	*	0.46	*	0.21	*	0.20	*		
8. DigitStdRatio	0.35	0.12	-0.31	*	-0.28	*	-0.20	*	-0.14	*	-0.14	*		
9. LocationStdRatio	0.39	0.15	-0.15		-0.17	*	-0.15		-0.12	*	0.12	*		
10. ShapeStdRatio	0.43	0.18	-0.21	*	-0.29	*	-0.30	*	-0.17	*	0.18	*		
11. DigitRunRatio	0.19	0.17	-0.29	*	-0.29	*	-0.18	*	-0.07	*	0.23	*	0.26	*
12. LocationRunRatio	0.29	0.25	-0.23	*	-0.23	*	-0.19	*	-0.25	*	0.22	*	0.17	*
13. ShapeRunRatio	0.37	0.24	-0.31	*	-0.31	*	-0.25	*	-0.09	*	0.23	*	0.25	*

* BF 3

‡ BF .33