Working Memory Capacity Development through Childhood: A Longitudinal Analysis

Matthew R. Reynolds,[a] Christopher R. Niileksela,[a] Gilles E. Gignac,[b] and Clarissa N. Sevillano[a]

The University of Kansas [a]

The University of Western Australia [b]

Submitted to *Developmental Psychology* on June 18, 2021
Resubmitted to *Developmental Psychology* on November 29, 2021
Accepted for publication in *Developmental Psychology*, February 14th, 2022.

Author Note

This research was supported by Texas Women's University Woodcock Institute Research grant. The data used in this study come from the U.S. Department of Education, National Center for Education Statistics, ECLS-K:2011. More information about the data and how to obtain them are located here: https://nces.ed.gov/ecls/datainformation2011.asp. This study was not preregistered.

Abstract

Working memory is an often studied and important psychological construct. The growth of working memory capacity (WMC) in childhood is described as linear. Average adult WMC is estimated as either 4 or 5 "chunks." Using latent curve models of data from a measure of digit span backward that was administered longitudinally to a large sample representative of the native English-speaking US kindergarten population in 2011, we found that WMC growth in childhood is curvilinear. It shows an increasing, yet decelerating pattern. Scoring rules (e.g., requiring 50% or 75% of trials correct) influence age-based estimates, but WMCs have likely been underestimated in children, and the average adult WMC of 5 is more plausible than 4, as measured by digit span backward. Developmental WMC estimates, such as those reported in this research, may help others develop prescriptive learning interventions for children and understand its growth and decline across the lifespan.

Keywords: digit span backward; working memory capacity; working memory development

Working Memory Capacity Development through Childhood: A Longitudinal Analysis

Working memory capacity (WMC) is the ability to maintain and manipulate information in active attention (Schneider & McGrew, 2019). It is critical to cognition and learning, and it correlates moderately with intelligence (Ackermann et al., 2005; Dempster, 1981; Fry & Hale, 1986; Gignac, 2014) and academic achievement (Peng et al., 2016; Peng et al., 2018). WMC is implicated in developmental disorders such as attention-deficit/hyperactivity disorder (ADHD; Ramos et al., 2020), specific learning disabilities (Swanson & Siegel, 2011), and intellectual disabilities (Bruns et al. 2019). Understanding its development is vital, and an accurate empirical description of WMC development through childhood is essential for informing theory (Best & Miller, 2010; Cowan, 2016) and specific educational practices (Cowan, 2014; Pickering, 2006). Current developmental descriptions are derived mostly from cross-sectional research with relatively small sample sizes. A longitudinal study with a large and representative sample followed through childhood would inform WMC development during this period.

WMC measurement

Digit span forward and backward are popular measures of memory (Rabin et al., 2016), and are important in the measurement of intelligence (e.g., Wechsler, 2014; Woodcock et al., 2001). On these tasks, an examiner reads sequences of single digits aloud at one-second intervals. An examinee repeats the sequences back in the same order on forward tasks, and in the reverse order on backward tasks. That simple response difference is critical (Ramsay & Reynolds, 1995). Scores from these two different tasks form distinct constructs when factor analyzed (Gignac et al., 2018; Reynolds, 1997), and correlate differently with other constructs—for instance, backward tasks correlate more strongly with general and fluid intelligence (Conway et al., 2002). Digit span forward measures short-term memory, whereas digit span backward measures WMC (McAuley & White, 2011; St. Clair-Thompson, 2010); the latter of which requires active manipulation of information while attending to a task (Baddeley, 1986; Engle & Kane, 2003; Gathercole et al., 2004; Gerton et al. 2004; Ramsay & Reynolds, 1995). WMC is the focus here.

WMC development during childhood

WMC development during childhood coincides with important biological processes such as maturation (Best & Miller, 2010; Tamnes et al., 2013) and environmental experiences such as formal schooling (de Wilde et al., 2016; Finch, 2019). Its growth during childhood is described as linear or constant (Best & Miller, 2010; Cowan, 2016; Gathercole et al., 2004). However, this description is based on relatively limited data, and other data suggest that growth may be steeper in early elementary school than in later elementary school, or curvilinear (Stipek & Valentino, 2015; Tulsky et al., 2013). The most accurate account possible is important, especially if growth in WMC is associated with other changes that also occur during this time (Courchesne et al. 2000; see Teffer & Semendeferi, 2012).

Cowan (2016) provided an exceptionally comprehensive review of why WMC improves over the course of development. To base his review on an empirical developmental pattern of WMC, he plotted the average number of items–including digits backward–recalled by children aged 4 to 15 years using one-year intervals. These estimates were based on means reported in cross-sectional data from what was viewed then as the most comprehensive developmental study of WMC in childhood and adolescence (Gathercole et al., 2004). According to the plot, six-year-old children, on average, recalled two digits in backward order, whereas 11-year-old children recalled slightly fewer than three digits in backward order. The growth trajectories were similar across most stimuli (e.g., digit span backward versus counting span), and were linear through childhood, leveling off in early adolescence. For example, digit span backward leveled off at about three digits for 13-, 14-, and 15-year-olds. Nonetheless, sample sizes were relatively small, ranging from 14 to 101 in each age group. In particular, after the age of 10 years no age group had more than 50 children in it.

Finch (2019) used longitudinal data from a large sample ($N > 11{,}000$) to study WMC growth from kindergarten through second grade with scores from a digit span backward (DSB) task, focusing on how schooling may affect WMC growth. WMC increased through second grade, but was slightly steeper through kindergarten. WMC growth during the school year was slightly faster than WMC growth during the summer. The limited developmental period precluded a description of WMC development across childhood, which we aimed to describe here.

Scoring methods and WMC estimates

One curious finding from Finch's (2019) study[1] was that by the end of second-grade, when students are likely about 7 or 8 years old, the average longest DSB was about 3.42 digits, or about a half digit more than the span of 13 to 15-year-olds in Cowan's (2016) study. Average WMC estimates from Finch's study and Cowan's study, as measured by DSB, differed substantially.

Digit span tasks typically include several trials for different digit sequence lengths. DSB therefore may be estimated in different ways. Finch's estimates were based on longest span correct; whereas Cowan's estimates were based on a study that required four or more out of six

[1] Growth analysis was based on Rasch-based ability scores. Longest digit sequences were not modeled but were reported descriptively.

trials to be correct. The requirement of four or more correct is stricter and may be the reason for the lower average DSB in Cowan's study. To reconcile the different findings, we aimed to estimate how different scoring procedures influence DSB averages and whether these scoring procedures affect estimates of growth trajectories.

Childhood WMC developmental estimates and the mystery of four or five

Developmental estimates of average WMC are discussed in adults, but not in children. Cowan (2001; 2010) described 4 digits (+/- 1) as the magical and mysterious average WMC obtained in adulthood. This capacity is in line with WMCs he reported in his empirical description of childhood WMC growth (Cowan, 2016). More recent researches, however, found that 5 digits (+/- 2.5) better describes adult WMC averages (Gignac, 2015; Gignac & Weiss, 2015).[2] In Finch's study, second-graders recalled 3.42 digits backward on average. Because it is unlikely that DSB increases by only one-half digit after second-grade, 5 digits seems like a more plausible estimate of average adult WMC based results from Finch's study. Regardless, more research is needed to confirm or disconfirm Gignac's (2015) finding of 5 being a more accurate average WMC (as measured by DSB) in adulthood. Hence, we aimed to a) produce developmental estimates of average WMC in childhood and b) use those data to further investigate the plausibility of either 4 or 5 average DSB in adulthood.

Study Purpose

Well-described empirical developmental definitions of constructs are critical for theory and research (Cowan, 2016; Dempster, 1981; Lykken, 1968). The current study aimed to better

[2] Gignac and Weiss's (2016) adult estimates were based on the longest DSB recalled. Each span length has two trials, and at least one trial needs to be correct to continue administration with longer spans.

understand and describe WMC development in childhood by addressing three gaps in the current literature.

First, the description of linear WMC development during childhood is based mostly on cross-sectional data from relatively small samples, and there is evidence to suspect that it may not be linear (Lensing & Elsner, 2018; Stipek & Valentino, 2015; Tulsky et al., 2013). A longitudinal study of WMC across childhood that compares different developmental trajectories in a large and representative sample of children should provide clarification. Additionally, to evaluate the generalizability of the developmental WMC trajectories, we analyzed the trajectories by sex due to different maturational patterns across childhood and adolescence (DeBellis et al., 2001; Simmonds et al., 2014).

Second, the average DSB at each age in childhood differs across studies (Cowan, 2016; Finch, 2019). However, studies have employed different scoring methods. Stricter scoring methods likely result in lower estimates of average DSB. It is not clear how much scoring differences affect these averages and if scoring differences account for different estimates of average DSB found across studies. Moreover, it is not clear how different scoring criteria affect estimated developmental trajectories. Research is needed to investigate different scoring procedures and their influence on WMC averages and trajectories that have been reported in research.

Third, WMCs at different ages have not been reported throughout childhood as they have in adulthood. Moreover, there are conflicting estimates of average adult WMC. Is a DSB of 4 (+/- 1) (Cowan, 2010) or 5 (+/- 2.5) (Gignac, 2015) a more plausible average? A comprehensive analysis of WMC, as measured by DSB, that produces developmental WMC estimates during childhood can be used to understand the rate at which it grows during childhood, and those estimates can be linked to research with adults to clarify conflicting estimates of average adult WMC.

We used three sets of questions to guide us in addressing three gaps in previous research:

1) Is average WMC growth, as measured by DSB, between ages 5 and 11 ½ years linear? Or is WMC growth, as measured by DSB, during childhood better described by a different trajectory? How much growth occurs between ages 5 to 11 ½? Are trajectories generalizable across sex?
2) Do different scoring procedures (e.g., longest DSB correct versus longest DSB associated with 50% of trials correct) explain different WMC estimates reported in other studies (Cowan, 2016; Finch, 2019)? Do different scoring procedures result in different developmental trajectories?
3) What are average developmental estimates for DSB during childhood? Are these estimates in childhood more consistent with an estimated average WMC of 4 or 5 in adulthood?

To answer these questions, we examined WMC trajectories based on DSB measured longitudinally in children from the ages of 5 to 11 ½ years, effectively doubling the time span reported in previous longitudinal research (Finch, 2019). The longitudinal data were from over 12,000 native English-speaking students who participated in the Early Childhood Longitudinal Study-Kindergarten: 2011 cohort (ECLS-K: 2011) study. Latent curve models were used to study trajectories. DSB estimates derived from different scoring methods were analyzed to investigate the influence of those different scoring methods on DSB estimates and their growth trajectories. Average DSB was calculated at different childhood ages and estimated with statistical models to provide developmental averages through childhood and to investigate the plausibility of the 4- or 5-digit WMC average in adulthood. In addition, percentages of 11-11 ½ year-old children who recalled 4 or more and 5 or more digits backward correct were calculated. What emerged was a comprehensive empirical description of WMC development in childhood.

Method

Sample

Data were from the ECLS-K: 2011 cohort. This nationally representative sample includes over 18,000 children from 968 different schools across the United States. Children were recruited in kindergarten, and then followed through 5th grade. There were nine measurement occasions, including the fall and spring of kindergarten, first, and second grades, and the spring of third, fourth, and fifth grades. Restricted-use data that included data on primary sampling units and child-level sample weights to account for oversampling and nonresponse at later time points were used. Only native English speakers were included in the analysis to minimize the influence of English language proficiency on scores (Ortiz, 2019). Thus, the total sample in this study included 12,330 participants. With sample weights applied, the sample was representative of the population of native English speaking kindergarten students in the United States when the study began. Of the participants, 48.0% were female and 52.0% were male. For ethnicity, 12.1% were Hispanic. For race, 64.0% were White, 15.6% were Black/African American, 1.6% were Asian, 0.4% were Native Hawaiian/Pacific Islander, 1.4% were Native American/Alaskan Native, and 4.8% reported two or more races. This research was reviewed by the IRB at the University of Kansas as part of a protocol for research with the ECLS-K:2011 and was determined to not be human subjects research because it used deidentified existing data (STUDY00143433, *Growth in Academic Skills for Children with Exceptionalities*).

Instrument

Numbers Reversed from the Woodcock Johnson III Tests of Cognitive Abilities (Woodcock et al., 2001) was administered as part of the ECLS-K: 2011 study. The measure includes 30 items. Items with 2- and 3-digit sequences (items 1-10) include 5 trials. Items with 4 through 8-digit sequences (items 11-30) include 4 trials. Trials were administered until three consecutive responses of the same digit sequence length were incorrect or until item 30 was reached (Tourangeau et al., 2019). For this study, the number of digits recalled backward on the last item the child answered correctly was used as the longest DSB for the initial analysis. These findings were compared with scoring that required at least 50% of the trials correct (50% on items 11-30, 60% on items 1-10) and at least 75% of the trials correct (75% on items 11-30, 80% on items 1-10).

Structuring Data

M*plus* 7.4 (Muthén & Muthén, 1998-20) was used for analyses. Stratification for primary sampling units and longitudinal weight variables were used to adjust for differential sampling rates and nonresponse across waves so sample estimates generalized to the United States population of kindergarteners in 2011.

Numbers Reversed was administered on up to nine occasions to each student from kindergarten through fifth grade. Rather than using the testing date or season of school (i.e., fall or spring), the age at which the child completed the test determined which column the datum for that child was included. Thirteen age “buckets” of 6-month intervals were created starting at 60-

65.99 months (i.e., 5 years 0 months to 5 years 6 months; 5:0-5:6)[3] and ending at 132.00-135.99 months (11 years 0 months to 11 years 6 months; 11:0-11:6).

Plan for Analysis

Univariate latent curve models (LCM) were used to model the data (Keith, 2019 Preacher et al., 2008). Each LCM included two common factors at a minimum: initial level and slope (i.e., trajectory). The initial DSB factor mean represented the model-implied average DSB at ages 5:0-5:6. The DSB slope factor mean for the linear model represented the model-implied average six-month increase in DSB across childhood. The two factor variances represented individual differences in initial DSB and in DSB slopes. Unstandardized factor loadings for the initial DSB factor were fixed to one. Unstandardized loadings for the slope factor(s) depended on the specified growth pattern: a) linear growth included loadings of 0, 1, 2, 3…12 across the equally spaced 6-month age "buckets", b) quadratic growth included the linear growth slope factor, and a second slope factor with loadings that corresponded to the linear growth factor loadings, but squared, and 3) empirically described growth (or latent basis model) included one slope factor with the first unstandardized slope factor loading fixed to 0 and the last loading fixed to 12–all other slope loadings were estimated freely (Preacher et al., 2008). Measured variable intercepts were fixed to zero, and their residual variances were set equal across all time points. Because of the complex survey and multilevel design, corrected standard errors and χ^2 tests of model fit were obtained by using a maximum likelihood estimator with robust standard errors.

[3] Analysis using age buckets with 3- and 4-month intervals was also conducted, but this analysis was not reported due to low covariance coverage that did not allow for corrected standard errors and χ^2. Nevertheless, based on uncorrected maximum likelihood estimates, both initial and slope factor estimates were similar across different the age-intervals.

Standalone fit indices were interpreted. Information criteria were used to compare competing models.

Data in this study were missing due to structuring the data with age “buckets” and attrition. We conducted Little’s missing completely at random (MCAR) test. This test was statistically significant, χ^2 (3803) = 4561.67, p < .01, suggesting missing data could not be assumed MCAR. Missing data were handled using full information maximum likelihood (FIML) estimation in M*plus* 7.4 and under the assumption that they were missing at random. Additionally, we conducted follow-up analysis to examine the influence of missingness and attrition on the latent intercept and slope (Diggle & Kenward, 1994). In this analysis, latent basis models were used with the original nine time points, representing different school semesters, and based on longest DSB correct. We created simple missing data variables for each time point (nine binary dummy-coded variables representing the presence or absence of data at each time point) and dropout variables representing attrition from the study (eight binary dummy-coded variables representing the time point at which participants dropped out of the study). Initial and slope factors were then regressed on the missing data and dropout dummy variables to examine missing data and attrition influences on initial level and slope estimates.

Last, the data used in this study come from the U.S. Department of Education, National Center for Education Statistics, ECLS-K:2011. More information about the data and how to obtain them are located here: https://nces.ed.gov/ecls/datainformation2011.asp. This study was not preregistered.

Results

Descriptives

Weighted mean estimates of the longest DSB sequence recalled in six-month time intervals are shown in Table 1 for the three different scoring procedures. DSB means increased across childhood, but six-month DSB mean increases were largest through age eight. The longest DSB correct scoring method showed the steepest average six month increases through age eight (M = .24); the 75-80% DSB correct scoring method showed the smallest six-month increases during that time (M = .19). After age eight, six-month DSB increases were similar across the scoring methods (Ms = .12-.13). As expected, the average DSB across the age range for the 75-80% DSB correct scoring (M = 2.62) was .72 less than the corresponding DSB average for the longest DSB correct scoring (M = 3.34). The mean difference between the average DSB across the age range for the 50-60% DSB correct scoring (M = 2.95) and longest DSB correct scoring was .39. Standard deviations for the longest DSB correct were similar across ages (M = .90). The other scoring methods showed increasing variability with age, but variability at the last measurement point was similar across scoring methods (SDs = .94-.99). All estimates were mostly normally distributed across ages.

The average correlations between scoring methods were all strong. The average correlation between longest DSB and 50-60% DSB correct was .83 (SD = .03, range = .78-.87); between longest DSB and 75-80% DSB correct was .69 (SD = .05, range = .64-.80); and between 50-60% DSB correct and 75-80% DSB correct was .79 (SD = .04, range = .75-.87).

Lastly, the percentages of 11-11½ year-olds who recalled at least four and five digits backward correctly were calculated across the three scoring methods. Using the longest DSB correct, 82.3% had a DSB of at least 4, and 33.0% of at least 5. Using the 50-60% DSB correct, 68.0% had a DSB of at least 4, and 21.3% of at least 5. Using 75-80% DSB correct, 47.9% had a DSB of at least 4, and 10.5% of at least 5.

Latent Curve Models

Model fit statistics for the LCMs are shown in Table 2. Linear LCMs fit poorly for the longest DSB estimates, although RMSEAs were acceptable. In general, linear models did not adequately describe WMC growth. Latent basis and quadratic LCM models showed, at minimum,"acceptable" fit statistics. According to BIC and aBIC, the quadratic LCM fit best for the longest DSB correct scoring. According to the AIC, the latent basis LCM fit best for the longest DSB correct scoring. Quadratic LCMs fit best for 50%-60% and 75%-80% correct scoring methods.

Model parameters are shown in Table 3. In all three quadratic LCMs, linear slope factor means were positive indicating average positive linear growth with age, whereas quadratic factor means were negative, indicating slowed growth with age. This same general pattern emerged in the latent basis LCMs, and this pattern is shown by the unstandardized factor loadings in Table 3. DSB increased across childhood, but the magnitude of change decreased with age. All factor variances were statistically significant—there were individual differences in initial levels of DSB, linear growth, and curvatures. Based on the standardized residuals, the quadratic models explained about 50% of the total standardized variance in WMC growth across the age range, but more variance was explained in the later time points than in the earlier time points when stricter scoring was used. Figure 1 shows quadratic model-implied trajectories for each scoring method.

The same modeling procedure was applied to subsamples divided by sex. Quadratic LCM models fit best. In those models, initial DSB means were slightly less for males (<.10 digit), but the linear and quadratic components of the slope factors were nearly identical. See supplemental tables and figures for these estimates.

Missing data analysis using dummy-coded missing data variables with the longest DSB correct data showed negligible to small effects on the latent intercept and slope. In combination, the dummy-coded missing data variables accounted for 1.5% of the variance in the latent intercept and 0.7% of the variance in the latent slope. Most individual effects were not statistically significant, small in magnitude, and were both positive and negative, effectively canceling each other out. Similar results were found with the dummy-coded dropout variables. These variables accounted for 1.3% of the variance in the latent intercept and 0.6% of the variance in the latent slope. Again, most individual effects were not statistically significant and small in magnitude. Missingness and attrition did not meaningfully predict or alter the study's results.

Discussion

We sought to provide a developmental description of WMC in native English-speaking children in the US based on a large and representative longitudinal sample using a measure of DSB. Additionally, we wanted to examine the influence of different scoring methods on WMC estimates and trajectories, and provide developmental estimates for WMC (based on DSB) across childhood. The developmental estimates were also used to inform the plausibility of either a 4 or 5 average WMC for adults.

WMC development during childhood

Using a measure of DSB, a curvilinear, increasing yet decelerating, growth pattern described WMC development from ages five through 11½ years. Although WMC growth during childhood is often described as linear (Best & Miller, 2010; Gathercole et al., 2004); the decelerating growth pattern found here is consistent with other data that have been suggestive of this pattern (Lensing & Elsner, 2018; Stipek & Valentino, 2015; Tulsky et al., 2013). These findings were consistent across scoring methods.

WMC doubled between the ages of 5 and 11½ years. Eleven-year-olds recalled, on average, about 1.9 to 2.3 more digits backward than five-year-olds across scoring methods. The WMC increase during this age range is more than double the increase expected for the rest of their lives. For instance, using the longest DSB correct scoring, the model implied average DSB for 11-11½-year-olds was 4.2, or about one less than the 5 DSB average reported for adults (Gignac, 2015; Gignac & Weiss, 2015)—if 4 DSB is used, adult WMC is already achieved, a finding that seems implausible (McArdle et al., 2002; McGrew et al., 2007). At a growth rate of .07 digit per six months, which is about the average six-month growth rate between ages 10 and 11½ years, average adult WMC of 5 would be reached by 17 years of age. Because the growth rate likely continues to decelerate beyond age 11, it may not be reached until later. Future research should map out the growth, peak, plateau, and decline of WMC across the lifespan. Cross-sectional data analysis shows that WMC peaks when adults are in their 20s and early 30s, and slowly decreases throughout adulthood (Grégoire & van der Linden, 1997; McArdle et al., 2002; McGrew et al., 2007).

Latent curve models were tested separately by sex due to differential maturation and to test the generalization of our findings. The findings were remarkably similar to each other and to the combined sample. Despite different maturational patterns in brain development (DeBellis et al., 2001), WMC trajectories (as measured by DSB) during childhood in this study are similar across the sexes.

Lastly, within-child WMC increases substantially during childhood, but there are also substantial WMC differences between children within each age range. Whereas the average increase in DSB was two digits between ages 5 and 11½, the DSB standard deviations within each age group were close to 1.[4] About 95% of children within each six-month age group are predicted to have a DSB +/- 2 digits from each age mean. Such individual differences in WMC are exceptionally important in learning (Peng et al., 2018; Swanson & Siegel, 2011), and highlight the importance of recognizing both intra-individual and individual differences.

Scoring methods and average DSB estimates

Different scoring methods led to different average DSB estimates in this study (cf. Cowan, 2016; Finch, 2019). Stricter scoring methods resulted in lower average DSB estimates, as expected. Scoring methods matter, but they do not fully account for differences in average DSB found across studies. Even using the strictest scoring criterion, the average DSB for 9-year olds in this study was about the same as 13- to 15-year-olds in Cowan's (2016) study.

At the age of six, estimates were not much different between our study and those reported by Cowan (2016). Lower DSB estimates at older age levels in childhood may be attributed to the way DSB was calculated by Cowan (2016). DSB estimates were derived from a Table presented in the Gathercole et al. (2004) study that reported mean DSB raw scores[5]; whereas we had direct DSB estimates for each child in our sample. This difference likely accounts for dissimilar findings because it is unlikely that there are large average differences in DSB estimates in US versus UK schoolchildren or substantial changes in WMC over time across cohorts (Gignac, 2015). The take home message from the current study is that regardless of how it is calculated,

[4] Some *SD*s for the 50-60% and 75-80% correct scoring were less than one at the younger ages.

[5] Mean DSB raw points were divided by six. This calculation assumes the examinee correctly answered every sequence within a trial correctly, although they needed only 4 out of 6 correct to move forward to another trial (if they answered the first 4 sequences correctly within a trial, full credit was awarded). Occasional misses within trials would lower the overall means.

average WMC in childhood appears to be larger than reported in previous research. Having access to raw scores from items is a strength of our dataset.

Last, although different scoring methods resulted in different DSB estimates, they did not alter the average trajectories much within this study. Average trajectories were slightly less curvilinear with stricter scoring conditions, but linear models did not adequately describe data well for any of the scoring methods.

Childhood WMC developmental estimates and the mystery of four or five

Developmental estimates through childhood were obtained with regard to DSB. According to the longest DSB correct scoring, on average, 5½ year-olds recall 2 digits backward, 7-year-olds recall 3 digits backward, and 10-year-olds recall 4 digits backward. In addition, about half or more of the 11-11½ year-olds in this study recalled at least 4 digits backward correct, regardless of the scoring. More than 80% of these students recalled 4 digits backwards at least one time. Taken together, unless there is no growth beyond age 11, these findings support recent findings of 5 being the magical number for WMC in adults (Gignac, 2015; Gignac & Weiss, 2015) rather than 4 (Cowan, 2001), at least as measured by DSB and in the United States.

Implications

One implication from the findings in this study is that WMC developmental trajectories seem to diverge from reading and mathematics developmental trajectories that increase rapidly throughout ages 5 to 18 (McGrew et al., 2007). Growth in WMC has been shown to correlate with growth in reading and mathematics skills during elementary school, but not beyond (Stipek & Valentino, 2015). It may be that accelerated WMC growth in early elementary school partially accounts for foundational reading and mathematics development, but the slowed, yet additional growth beyond that time does not account for more complex learning in those areas (e.g., Hajovsky et al., 2014).

Instructional strategies may be improved by prescribing them based on their working memory demands (Cowan, 2014; Pickering, 2006). Developmental estimates based on WMC as described in this study, rather than age standardized scores, may be useful in developing and understanding prescriptive learning strategies. It may be that certain WMCs are optimal for specific teaching approaches and learning strategies, or for the amount of new information that is presented. Intervention strategies that have reduced WMC demands may be more appropriate for students with smaller WMC and may be more effective by attending to these individual differences.

Limitations

Other measures of WMC with different stimuli or content have been found to follow similar trajectories (e.g., Gathercole et al., 2004; Mathy & Freidman, 2020). Nevertheless, findings from this study are limited to DSB. Different trajectories and developmental estimates may be found when measures with different stimuli are used in very large samples.

The findings from this study are representative of the native English-speaking US kindergarten population in 2011. Initial means and trajectories represent averages and may not generalize to different cohorts, different nations, or every possible subgroup within this study (e.g., in this study females had slightly higher initial means than males). The average trajectory and the "magical" average WMC found in this study and described elsewhere (Cowan, 2001) should not be assumed to apply universally.

Children in this sample were administered the same test up to nine times, opening up the possibility of practice effects. Research shows practice effects with frequent assessments over the course of weeks (Bartels et al., 2010), but at least with adults, consistent effects are less clear when it is over many months or years (Salthouse, 2010). Although we cannot eliminate the possibility of practice effects, we have evidence to suggest that they did not substantially influence the findings in this research.

First, age-based standard scores (i.e., $M = 100$, $SD = 15$) for the Numbers Reversed test in our sample were calculated for each test administration. If DSB performance improved because of practice effects, the standard scores should increase over time because the participants in our study were exposed to the test multiple times whereas the participants in the standardization sample were exposed to the test once. All standard scores were between 96.0 and 99.5 (M = 96.8). There was no increase over time. The highest standard score was at the first administration of the test. Thus, the results reported here are unlikely to have been influenced by practice effects.

Second, we located age-based cohort DSB estimates from two versions of the Wechsler Intelligence Scales for Children (WISC), the WISC-IV (Wechsler, 2004) and WISC-V (Wechsler, 2014). Average longest DSB estimates are reported by year of age, starting at 6 years. Those estimates are based on $n = 200$ in each age year sample, and represent the longest DSB recalled at least once (out of two administered trials). Compared to our longest DSB estimates, the average longest DSBs on the WISCs were consistent with our model-implied DSB averages at ages six (2.6 DSB; WISC-IV/V = 2.6/2.7 DSB) and eleven years (4.2 DSB; WISC-IV/V = 4.1/4.2 DSB). These findings provide additional support that practice effects were unlikely in this investigation. Moreover, this converging evidence from our longitudinal sample and two nationally representative cross-sectional samples (based on different tests developed 10 years apart) bolsters the finding that WMC in childhood is larger than previously estimated.

Conclusion

The study contributed to understanding WMC development during childhood, as measured by DSB in five ways. First, WMC more than doubles in size from ages 5 to 11½ years, although within each age level there are also large individual differences. Second, WMC has an increasing, yet decelerating growth pattern through childhood. Third, scoring procedures used in the calculation of WMC influence average WMCs reported in studies, but not so much the trajectories. Fourth, despite different estimates across scoring methods, average WMC estimates appear to have been underestimated in childhood. Lastly, an average adult WMC of 5 (Gignac, 2015) seems more plausible than 4 (Cowan, 2010).

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*, 30-60. https://doi.org/10.1037/0033-2909.131.1.30

Baddeley A. (1986). *Working memory*. Oxford University Press, Oxford.

Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., & Ehrenreich, H. (2010). Practice effects in healthy adults: A longitudinal study on frequent repetitive cognitive testing. *BMC Neuroscience. 11*, 118. https://doi.org/10.1186/1471-2202-11-118

Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 National Survey. *Journal of School Psychology*, *72*, 29-48. https://doi.org/10.1016/j.jsp.2018.12.004

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*, 1641-1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Bruns, G., Ehl, B., & Grosche, M. (2019). Verbal working memory processes in students with mild and borderline intellectual disabilities: Differential developmental trajectories for rehearsal and reintegration. *Frontiers in Psychology*, *9*, 2581. https://doi.org/10.3389/fpsyg.2018.02581

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163-183. https://doi.org/10.1016/S0160-2896(01)00096-4

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioural and Brain Sciences, 24*, 87–185. doi:10.1017/S0140525X01003922

Courchesne, E., Chisum, H. J., Townsend, J., Cowles, A., Covington, J., Egaas, B., ... & Press, G. A. (2000). Normal brain development and aging: quantitative analysis at in vivo MR imaging in healthy volunteers. Radiology, 216(3), 672-682. https://doi.org/10.1148/radiology.216.3.r00au37672

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*, 51–57. https://doi.org/10.1177/0963721409359277

Cowan N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, *26*, 197–223. https://doi.org/10.1007/s10648-013-9246-y

Cowan, N. (2016). Working memory maturation: Can we get at the essence of cognitive growth? *Perspectives on Psychological Science*, *11*, 239-264. https://doi.org/10.1177/1745691615621279

De Bellis, M. D., Keshavan, M. S., Beers, S. R., Hall, J., Frustaci, K., Masalehdan, A., ... & Boring, A. M. (2001). Sex differences in brain maturation during childhood and adolescence. Cerebral Cortex, 11(6), 552-557. https://doi.org/10.1093/cercor/11.6.552

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin*, *89*, 63-100. https://doi.org/10.1037/0033-2909.89.1.63

de Wilde, A., Koot, H. M., & van Lier, P. A. (2016). Developmental links between children's working memory and their social relations with teachers and peers in the early school years. *Journal of Abnormal Child Psychology*, *44*, 19-30. https://doi.org/10.1007/s10802-015-0053-4

Diggle, P., & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *43*(1), 49-73. https://doi.org/10.2307/2986113

Engle, R. W., & Kane, M. J. (2004). Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory,* Vol. 44, pp. 145–199). Elsevier Science.

Finch, J. E. (2019). Do schools promote executive functions? Differential working memory growth across school-year and summer months. *AERA Open*, *5*, https://doi.org/10.1177/2332858419848443.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*, 237-241. https://doi.org/10.1111/j.1467-9280.1996.tb00366.x

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory from 4 to 15 Years of Age. *Developmental Psychology*, *40*, 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gerton, B. K., Brown, T. T., Meyer-Lindenberg, A., Kohn, P., Holt, J. L., Olsen, R. K., & Berman, K. F. (2004). Shared and distinct neurophysiological components of the digits forward and backward tasks as revealed by functional neuroimaging. *Neuropsychologia*, *42*, 1781-1787. https://doi.org/10.1016/j.neuropsychologia.2004.04.023

Gignac, G. E. (2014). Fluid intelligence shares closer to 60% of its variance with working memory capacity and is a better indicator of general intelligence. *Intelligence*, *47*, 122-133. https://doi.org/10.1016/j.intell.2014.09.004

Gignac, G. E. (2015). The magical numbers 7 and 4 are resistant to the Flynn effect: No evidence for increases in forward or backward recall across 85 years of data. *Intelligence, 48*, 85-95. https://doi.org/10.1016/j.intell.2014.11.001

Gignac, G. E., & Weiss, L. G. (2015). Digit Span is (mostly) related linearly to general intelligence: Every extra bit of span counts. *Psychological Assessment*, *27*, 1312-1323. https://doi.org/10.1037/pas0000105

Gignac, G. E., Kovacs, K., & Reynolds, M. R. (2018). Backward and forward serial recall across modalities: An individual differences perspective. *Personality and Individual Differences*, *121*, 147-151. https://doi.org/10.1016/j.paid.2017.09.033

Grégoire, J., & van der Linden, M. (1997). Effect of age on forward and backward digit spans. *Aging, Neuropsychology, and Cognition*, *4*, 140-149. https://doi.org/10.1080/13825589708256642

Hajovsky, D., Reynolds, M. R., Floyd, R. G., Turek, J. J., & Keith, T. Z. (2014). A multigroup investigation of latent cognitive abilities and reading achievement relations. *School Psychology Review, 43*, 385-406. https://doi.org/10.1080/02796015.2014.12087412

Institute for Educational Sciences National Center for Education Statistics (IES NCES) (2019). Early Childhood Longitudinal Study, Kindergarten Class of 2010-11: Kindergarten-Fifth Grade Restricted Data (NCES 2019100) [Data set]. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2019100

Keith, T. Z. (2019). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (3rd ed.), Routledge, New York, NY.

Lensing, N., & Elsner, B. (2018). Development of hot and cool executive functions in middle childhood: three-year growth curves of decision making and working memory updating. *Journal of Experimental Child Psychology*, *173*, 187-204. https://doi.org/10.1016/j.jecp.2018.04.002

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, *70*, 151-159. https://doi.org/10.1037/h0026141

Mathy, F., & Friedman, O. (2020). Working memory develops at a similar rate across diverse stimuli. *Journal of Experimental Child Psychology*, *191*, 104735. https://doi.org/10.1016/j.jecp.2019.104735

McArdle, J. J., Ferrer-Caja, E., Hamagami, F. & Woodcock, R. W. (2002). Comparative longitudinal structural analysis of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology, 38, 1*, 115-142. https://doi.org/10.1037/0012-1649.38.1.115

McAuley, T., & White, D. A. (2011). A latent variables examination of processing speed, response inhibition, and working memory during typical development. *Journal of experimental Child Psychology, 108*, 453-468. https://doi.org/10.1016/j.jecp.2010.08.009

McGrew, K. S., & Schrank, F. A., & Woodcock, R. W. (2007). Technical manual. Woodcock-Johnson III normative update. *Rolling Meadows, IL: Riverside*.

Muthén, L. K., & Muthén, B. O. (1998-2012). Mplus User's Guide. Seven Edition. Los Angeles, CA: Muthén & Muthén.

Ortiz, S. O. (2019). On the measurement of cognitive abilities in English learners. *Contemporary School Psychology*, *23*, 68-86. https://doi.org/10.1007/s40688-018-0208-8

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*, *108*, 455-473. https://doi.org/10.1037/edu0000079

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., ... & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, *144*, 48-76. http://dx.doi.org/10.1037/bul0000124

Pickering, S. J. (2006). Working memory in dyslexia. In T. P. Alloway & S. E. Gathercole (Eds.), *Working memory and neurodevelopmental disorders* (p. 7–40). Psychology Press.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. SAGE Publications, Inc.

Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology, 31*, 206–230. https://doi.org/10.1093/arclin/acw007

Ramos, A. A., Hamdan, A.C., & Machado, L. (2020). A meta-analysis on verbal working memory in children and adolescents with ADHD. *The Clinical Neuropsychologist*, *34*, 873-898. https://doi.org/10.1080/13854046.2019.1604998

Ramsay, M. C., & Reynolds, C. R. (1995). Separate digits tests: A brief history, a literature review, and a reexamination of the factor structure of the Test of Memory and Learning (TOMAL). *Neuropsychology Review*, *5*, 151-171. https://doi.org/10.1007/BF02214760

Reynolds, C. R. (1997). Forward and backward memory span should not be combined for clinical analysis. *Archives of Clinical Neuropsychology, 12*, 29-40. https://doi.org/10.1016/S0887-6177(96)00015-7

Salthouse, T. A. (2010). Influence of age on practice effects in longitudinal neurocognitive change. *Neuropsychology, 24*, 563-572. https://doi.org/10.1037/a0019026

Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theory, tests, and issues* (4th ed., pp. 73–163). New York: Guilford.

Simmonds, D. J., Hallquist, M. N., Asato, M., & Luna, B. (2014). Developmental stages and sex differences of white matter and behavioral development through adolescence: a longitudinal diffusion tensor imaging (DTI) study. Neuroimage, 92, 356-368. https://doi.org/10.1016/j.neuroimage.2013.12.044

Sowell, E. R., Delis, D., Stiles, J., & Jernigan, T. L. (2001). Improved memory functioning and frontal lobe maturation between childhood and adolescence: A structural MRI study. *Journal of International Neuropsychological Society, 7*, 312–322. https://doi.org/10.1017/s135561770173305x

Stipek, D., & Valentino, R. A. (2015). Early childhood memory and attention as predictors of academic growth trajectories. *Journal of Educational Psychology*, *107*, 771–788. https://doi.org/10.1037/edu0000004

Swanson, H. L., & Siegel, L. (2011). Learning disabilities as a working memory deficit. *Experimental Psychology, 49*, 5-28.

St Clair-Thompson, H. L. (2010). Backwards digit recall: A measure of short-term memory or working memory? *European Journal of Cognitive Psychology*, *22*, 286-296. https://doi.org/10.1080/09541440902771299

Tamnes, C. K., Walhovd, K. B., Grydeland, H., Holland, D., Østby, Y., Dale, A. M., & Fjell, A. M. (2013). Longitudinal working memory development is related to structural maturation of frontal and parietal cortices. *Journal of Cognitive Neuroscience*, *25*, 1611-1623. https://doi.org/10.1162/jocn_a_00434

Teffer, K., & Semendeferi, K. (2012). Human prefrontal cortex: evolution, development, and pathology. Progress in Brain Research, 195, 191-218. https://doi.org/10.1016/B978-0-444-53860-4.00009-X

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., Hagedorn, M. C., Daly, P., & Najarian, M. (2015). Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K: 2011). User's Manual for the ECLS-K: 2011 Kindergarten Data File and Electronic Codebook, Public Version. NCES 2015-074. *National Center for Education Statistics*.

Tulsky, D. S., Carlozzi, N. E., Chevalier, N., Espy, K. A., Beaumont, J. L., & Mungas, D. (2013). NIH toolbox cognition battery (CB): measuring working memory. *Monographs of the Society for Research in Child Development, 78, 70-87.* https://doi.org/10.1111/mono.12035

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th ed.). San Antonio, TX: Psychological Corporation.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed.). Bloomington, MN: Pearson.

Woodcock, R. W., Mather, N., McGrew, K. S., & Wendling, B. J. (2001). *Woodcock-Johnson III tests of cognitive abilities*. Itasca, IL: Riverside Publishing Company.

Table 1

Descriptive Statistics

Age	*N*	*Longest Digit Span Backward Correct*			*Longest Digit Span Backward with 50-60% Trials Correct*			*Longest Digit Span Backward with 75-80% Trials Correct*		
		M (*SD*)	Skew.	Kurt.	*M* (*SD*)	Skew.	Kurt.	*M* (*SD*)	Skew.	Kurt.
5.0-5.5	4,150	1.96 (0.91)	0.48	-0.61	1.69 (0.70)	0.81	0.96	1.56 (0.60)	0.71	0.66
5.6-5.11	9,020	2.26 (0.93)	0.05	-0.86	1.92 (0.75)	0.45	0.04	1.75 (0.64)	0.48	0.77
6.0-6.5	8,210	2.57 (0.92)	-0.08	0.26	2.17 (0.79)	0.43	0.76	1.96 (0.67)	0.60	2.49
6.6-6.11	7,140	2.90 (0.89)	-0.24	0.45	2.48 (0.82)	0.20	0.11	2.21 (0.71)	0.39	0.62
7.0-7.5	6,480	3.11 (0.83)	-0.30	0.56	2.67 (0.82)	0.18	-0.12	2.35 (0.71)	0.49	0.34
7.6-7.11	6,140	3.32 (0.85)	-0.11	1.20	2.89 (0.85)	0.19	0.35	2.53 (0.75)	0.55	0.62
8.0-8.5	5,010	3.47 (0.84)	0.12	1.18	3.03 (0.86)	0.17	-0.11	2.65 (0.79)	0.54	-0.01
8.6-8.11	4,860	3.65 (0.87)	0.07	1.38	3.25 (0.89)	0.13	0.20	2.86 (0.83)	0.42	0.29
9.0-9.5	4,260	3.77 (0.89)	0.17	1.19	3.37 (0.91)	0.21	0.57	2.96 (0.86)	0.40	0.27
9.6-9.11	4,250	3.97 (0.93)	0.39	1.34	3.57 (0.93)	0.22	0.86	3.15 (0.89)	0.30	0.04
10.0-10.5	3,960	4.04 (0.92)	0.43	1.57	3.64 (0.93)	0.20	0.76	3.22 (0.89)	0.28	0.21
10.6-10.11	3,930	4.19 (0.99)	0.61	1.56	3.82 (0.97)	0.33	1.06	3.36 (0.93)	0.40	0.68
11.0-11.5	3,550	4.27 (0.99)	0.45	1.08	3.88 (0.99)	0.27	0.84	3.45 (0.94)	0.27	0.32

Note. Skew. = Skewness, Kurt. = Kurtosis. *N*s are rounded to the nearest 10 per requirements for using the restricted data for the ECLS-K: 2011. Ages are in years and months and represent a range. All descriptive statistics represent values using sample weights. *N*s using sample weights ranged from 1,004,423 to 2,184,560.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011), fall 2010, spring 2011, fall 2011, spring 2012, fall 2012, spring 2013, spring 2014, spring 2015, spring 2016.

Table 2

Latent Curve Model Fit Statistics

	S-B χ^2 (*df*)	CFI	TLI	RMSEA	SRMR	AIC	BIC	aBIC
Longest Digit Span Backward Correct								
1. Linear	1608.10 (98)	.835	.868	.035	.106	214938	214982	214963
2. Quadratic	334.81 (94)	.974	.978	.014	.050	211068	**211142**	**211110**
3. Latent Basis	327.22 (87)	.974	.976	.015	.065	**211058**	211184	211130
Longest Digit Span Backward with 50-60% Trials Correct								
4. Linear	1120.31 (98)	.890	.912	.029	.055	202141	202185	202166
5. Quadratic	315.41 (94)	.976	.980	.014	.054	**199686**	**199760**	**199729**
6. Latent Basis	370.28 (87)	.969	.973	.016	.060	199868	199994	199940
Longest Digit Span Backward with 75-80% Trials Correct								
7. Linear	888.15 (98)	.901	.921	.026	.067	185789	185834	185815
8. Quadratic	457.82 (94)	.954	.962	.018	.069	**184449**	**184523**	**184491**
9. Latent Basis	527.46 (87)	.945	.951	.020	.074	184676	184802	184748

Note. S-B = Satorra-Bentler, CFI = comparative fit index, TLI = Tucker-Lewis index, RMSEA = root mean square error of approximation, SRMR = standardized root mean square residual, AIC = Akaike information criterion, BIC = Bayesian information criterion, aBIC = sample-size adjusted Bayesian information criterion. Bolded information criteria indicate the best fitting model.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011), fall 2010, spring 2011, fall 2011, spring 2012, fall 2012, spring 2013, spring 2014, spring 2015, spring 2016.

Table 3

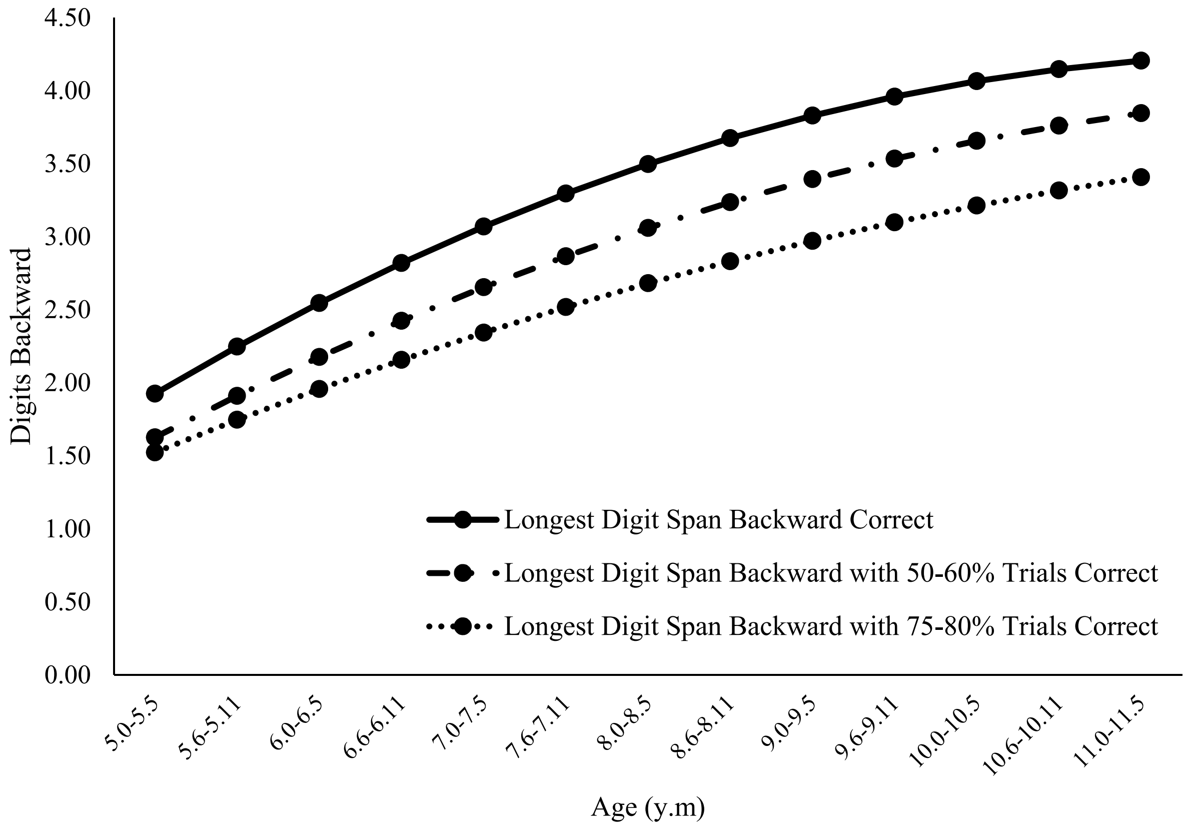
Latent Curve Model Parameter Estimates

	Longest Digit Span Backward Correct			*Longest Digit Span Backward with 50-60% Trials Correct*			*Longest Digit Span Backward with 75-80% Trials Correct*		
	Linear Model	Quad. Model	Latent Basis Model	Linear Model	Quad. Model	Latent Basis Model	Linear Model	Quad. Model	Latent Basis Model
Latent Intercept									
M	2.21	1.93	1.86	1.83	1.63	1.63	1.66	1.52	1.53
(*SD*)	(0.60)	(0.74)	(0.68)	(0.48)	(0.49)	(0.50)	(0.36)	(0.36)	(0.37)
Latent Slope									
M	0.19	0.33	0.20	0.19	0.29	0.19	0.16	0.23	0.16
(*SD*)	(0.05)	(0.15)	(0.06)	(0.05)	(0.11)	(0.05)	(0.05)	(0.10)	(0.05)
Quadratic Slope									
M	--	-0.01	--	--	-0.01	--	--	-0.01	--
(*SD*)		<(0.01)			<(0.01)			<(0.01)	
Correlations									
Int. & Slope	-0.35	-0.65	-0.54	-0.06	-0.20	-0.19	0.10	-0.08	-0.01
Int. & Quad. Slope	--	0.59	--	--	0.16	--	--	0.15	--
Slope & Quad. Slope	--	-0.92	--	--	-0.87	--	--	-0.85	--
Unstandardized Slope Loadings	Linear	Linear/ Quadratic	Latent Basis	Linear	Linear/ Quadratic	Latent Basis	Linear	Linear/ Quadratic	Latent Basis
5.0-5.5	0	0/0	0	0	0/0	0	0	0/0	0.00
5.6-5.11	1	1/1	1.72	1	1/1	1.38	1	1/1	1.26
6.0-6.5	2	2/4	3.53	2	2/4	2.86	2	2/4	2.64
6.6-6.11	3	3/9	5.04	3	3/9	4.39	3	3/9	4.05
7.0-7.5	4	4/16	6.23	4	4/16	5.50	4	4/16	5.08
7.6-7.11	5	5/25	7.14	5	5/25	6.55	5	5/25	6.03
8.0-8.5	6	6/36	8.07	6	6/36	7.54	6	6/36	7.05
8.6-8.11	7	7/49	8.80	7	7/49	8.43	7	7/49	8.09
9.0-9.5	8	8/64	9.53	8	8/64	9.29	8	8/64	8.94
9.6-9.11	9	9/81	10.39	9	9/81	10.08	9	9/81	9.84

	Longest Digit Span Backward Correct			*Longest Digit Span Backward with 50-60% Trials Correct*			*Longest Digit Span Backward with 75-80% Trials Correct*		
	Linear Model	Quad. Model	Latent Basis Model	Linear Model	Quad. Model	Latent Basis Model	Linear Model	Quad. Model	Latent Basis Model
10.0-10.5	10	10/100	10.81	10	10/100	10.72	10	10/100	10.50
10.6-10.11	11	11/121	11.42	11	11/121	11.38	11	11/121	11.10
11.0-11.5	12	12/144	12.00	12	12/144	12.00	12	12/144	12.00

Note. All linear and quadratic slope loadings were fixed. For the quadratic LCM loadings, the first number is the loading on the linear slope factor, the second number is the loading on the quadratic slope factor. In the latent basis model, only the first and last slope loadings are fixed, the others are freely estimated.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011), fall 2010, spring 2011, fall 2011, spring 2012, fall 2012, spring 2013, spring 2014, spring 2015, spring 2016.

Figure 1



Quadratic model-implied digit span backward trajectories for three scoring methods.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011), fall 2010, spring 2011, fall 2011, spring 2012, fall 2012, spring 2013, spring 2014, spring 2015, spring 2016.