Journal of Experimental Child Psychology 104 (2009) 252–265

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Investigating the childhood development of working memory using sentences: New evidence for the growth of chunk capacity

Amanda L. Gilchrist, Nelson Cowan *, Moshe Naveh-Benjamin

Department of Psychological Sciences, University of Missouri, Columbia, MO 65211, USA

ARTICLE INFO

Article history:
Received 20 November 2008
Revised 13 May 2009
Available online 17 June 2009

Keywords:
Child development
Sentence memory
Chunking
Working memory
Capacity limits
Verbal memory

ABSTRACT

Child development is accompanied by a robust increase in immediate memory. This may be due to either an increase in the number of items (chunks) that can be maintained in working memory or an increase in the size of those chunks. We tested these hypotheses by presenting younger and older children (7 and 12 years of age) and adults with different types of lists of auditory sentences: four short sentences, eight short sentences, four long sentences, and four random word lists, each read with a sentence-like intonation. Young children accessed (recalled words from) fewer clauses than did older children or adults, but no age differences were found in the proportion of words recalled from accessed clauses. We argue that the developmental increase in memory span was due to a growing number of chunks present in working memory with little role of chunk size.

Introduction

One can examine developmentally two different aspects of working memory, the small amount of information temporarily held in mind at any time. One can examine the number of separate units, or chunks, that are maintained (e.g., Broadbent, 1975; Cowan, 2001; Miller, 1956) as well as the size of each chunk. This chunk size depends on the participant's knowledge so that, for example, a chunk can be either a familiar word presented in isolation or a well-learned word pair (Chen & Cowan, 2005, 2009; Cowan, Chen, & Rouder, 2004). Previous research documents a robust increase in working memory capacity as children develop (see Cowan & Alloway, 2009, for a review). Is this due to an increase

* Corresponding author.
E-mail address: cowann@missouri.edu (N. Cowan).

0022-0965/$ - see front matter
doi:10.1016/j.jecp.2009.05.006

in the number of chunks that can be maintained, the sizes of chunks, or both? There is long-standing debate regarding this question.

Gilchrist, Cowan, and Naveh-Benjamin (2008) used the coherence of words within clauses to examine working memory decline in adult aging. They found that aging was accompanied by a decline in the number of chunks that could be recalled, with no apparent decline in chunk size. Specifically, in memory for lists of spoken sentences, older adults recalled words from fewer clauses but recalled the same amount as young adults from any clause that was at least partly recalled. This was referred to as diminished clause *access* with undiminished clause *completion*. This was interpreted as an aging decline in chunk capacity, with no change for these materials in chunk size. Here we examine whether the developmental difference between children and adults is similar in nature. To set the stage, we briefly review the empirical basis for a chunk limit in working memory, discuss competing theories of developmental increases in capacity, and discuss possible developmental influences on the formation and retention of chunks from linguistic materials.

Chunk capacity and chunk size in working memory

Philosophers and early experimental psychologists proposed a limited working memory (e.g., James, 1890) leading to limits in immediate recall of approximately seven items (e.g., Miller, 1956). Other research has shown, however, that other factors affect the amount that can be recalled. Baddeley, Thomson, and Buchanan (1975) showed that adults typically could recall as much as they could pronounce in approximately 2 s in serial recall; this time limit was based on refreshment of a temporary representation through covert rehearsal. The ''magical number seven'' reported by Miller (1956) could reflect the typical number of monosyllabic words recalled when a basic capacity was supplemented by rehearsal or even on-line formation of multi-item chunks. Subsequent work suggested that when the supplementary processes are eliminated, the limit in adults is in the range of three to five chunks (Broadbent, 1975; Cowan, 2001).

Given the 2-s limit for a phonological rehearsal loop (Baddeley, 1986; Baddeley et al., 1975), one way to examine the chunk limit is to use chunks so long that they preclude use of the loop, as in coherent linguistic materials. In the earliest such study, Tulving and Patkau (1962) presented series of words that varied in their approximation to English text, ranging from nonsensical to fully coherent. Free recall was examined for ''adopted chunks,'' sets of words recalled in their correct serial order. The authors found an invariant limit in recall of approximately four to six adopted chunks for all materials, but the size of adopted chunks (number of words per chunk) was dependent on the level of approximation to English. Subsequent studies with linguistic materials (Glanzer & Razel, 1974; Simon, 1974) suggest slightly lower limits of three or four linguistic chunks on average, more in keeping with Broadbent (1975) and Cowan (2001).

Developmental increases in memory capacity

Some research suggests a developmental increase in the number of chunks that can be maintained. Pascual-Leone (1970, 2005) proposed that as children age, the number of available items that can be maintained increases by a specified invariant number of age-related units. It was claimed that performance increases with age could not be explained by the development of strategic factors, such as the ability to form chunks, or by increases in knowledge (see Dempster, 1981, for a review). In one experiment, for example, Pascual-Leone (1970) taught different stimulus–response relations (e.g., raise your hand for a square, clap your hands for a red item) and then combined the cues into multidimensional signals (e.g., raise your hand and also clap your hands for a red square). The number of dimensions that children could handle increased with age. Working in this tradition, Burtis (1982) provided sets of items that could form multi-item chunks (e.g., repetitions of the same letter, both letters in a pair colored red, pairs familiar through repetition or forming known acronyms) and found an increase in the number of chunks recalled. Similar results have been obtained in procedures where each presented item is assumed to comprise a separate chunk because there is not enough free processing time devoted to the stimuli to allow them to be rehearsed or grouped (Cowan et al., 2005; Cowan, Naveh-Benjamin, Kilb, & Saults, 2006; Cowan, Nugent, Elliott, Ponomarev, & Saults, 1999).

Based on other studies, however, there are also reasons to expect that the average size of a chunk could increase with age. As children grow, they gain experiential knowledge and sophisticated linguistic ability (Chi, 1978; Ottem, Lian, & Karlsen, 2007). Also, children begin to use strategies to guide memory (Case, Kurland, & Goldberg, 1982; Flavell, Beach, & Chinsky, 1966; Harris & Burke, 1972; Kail, 1992; Ornstein, Naus, & Liberty, 1975; Towse, Hitch, & Skeates, 1999). As these increase with development, it is assumed that the size of chunks will also increase because these can increase item meaningfulness. Recently, Ottem and colleagues (2007) proposed that the characteristic developmental increase in memory span was due solely to growth in linguistic abilities. Intuitively, language allows a person to better rehearse and encode items (Flavell et al., 1966) as well as to form associations that aid chunking of items, proposed to be a critical factor for the span increase. In contrast to Pascual-Leone and other neo-Piagetians, Ottem and colleagues (2007) maintained that it was the size, not the number, of available chunks that increased with age.

Given this controversy, it is difficult to know whether the number of chunks in working memory, as opposed to chunk size, actually does change with age. We would be in an especially advantageous situation if chunk size remained constant across age groups, whereas the total amount recalled changed. That pattern of results could unequivocally be interpreted as a change in capacity in chunks. That was the result obtained by Gilchrist and colleagues (2008) for adult aging using lists of simple unrelated sentences to be recalled. Although the finding could be similar for child development, there are considerations of language development that theoretically could lead to a different outcome.

Controlling developmental differences in processing language stimuli

Some aspects of language improve throughout childhood (e.g., Chomsky, 1969). For our purposes, linguistic stimuli must be chosen with the aim of minimizing the differences as they pertain to working memory. There could be a developmental increase in the ability to use relatively simple linguistic structure to remember the details of short sentences or in the ability to use more complex linguistic structure to remember the two halves of long sentences. Young children's processing of the structure under a working memory load might suffer from the unavailability of resources needed to carry out the linguistic processing, as suggested by some investigators (Daneman & Case, 1981; Kail & Hall, 2001). Reduced processing efficiency may correspond to less complete syntactic or semantic encoding of language in younger children. Consequently, what is a single chunk to an older participant could be encoded as multiple chunks in a young child.

Hopefully, however, the simplicity of stimulus materials will minimize such differences. The basic retention of sentences falling well within the individual's linguistic competence may occur automatically with few demands on attention (Allen & Baddeley, 2009; Caplan, Waters, & DeDe, 2007). If so, we can anticipate that participants in all age groups will have similar chunk sizes for the materials and will differ only in the number of unrelated linguistic units or chunks held in memory.

The current study

Our primary goal was to observe age differences in the number of chunks held in working memory, but phonological storage and rehearsal could make a contribution to recall along with a chunk-capacity-limited mechanism. Chen and Cowan (2005) found a greater contribution of phonological processes for serial recall as opposed to free recall or free scoring of serial recall. Therefore, we were able to minimize the contribution of phonological rehearsal processes not only by using lists of sentences so as to greatly exceed the 2-s limit of phonological rehearsal in adults (e.g., Baddeley et al., 1975) but also by requiring free recall rather than serial recall.

We varied the number of unrelated sentences within a list as well as their length. To examine the development of this capacity, we presented children with the following: (a) lists of four short sentences, each with one independent clause; (b) lists of eight such short sentences; (c) lists of four long sentences, each composed of two meaningfully conjoined clauses; and (d) lists of four random pseudo-sentences made up of various words mixed together in a haphazard manner. We counterbalanced materials so that participants never received the same stimulus throughout the experiment. Every

Table 1
Illustration of the manipulation of stimulus examples across groups.

Group 1 stimulus example	Group 2 version of example
Our neighbor sells vegetables but he also makes fruit juice (4 long)	Our neighbor sells vegetables He also makes fruit juice (4, 8 short)
A close football your cheese (4 random)	Flag football starts soon She prepared a cheese sauce (4, 8 short)
Take your paper and pencil (4, 8 short)	Take your paper and pencil and answer every question (4 long) OR Lightning paper we bees take (4 random)

Note. Both groups received the same stimulus content words but in different conditions to balance materials across conditions.

long sentence could be broken down into two comprising short sentences; participants in one group received a long sentence, and those in the other group received its comprising sentences (see Table 1).

On the basis of these conditions, we could observe whether there are age differences in the access and completion of clauses, indexing the number and size of chunks recalled, respectively. The four short sentences condition was included to allow a sensitive age comparison given previous research suggesting that most young adults have an immediate capacity limit of approximately four chunks (e.g., Chen & Cowan, 2005; Cowan, 2001). The list conditions, summarized in Table 2, permit several critical comparisons. First, the lists of four short and eight short sentences differ in list length but not in the amount of coherence within a sentence. If there is a constant capacity mechanism and no effect of list length in this procedure, the result should be access to the same number of sentences in both conditions. This should differ according to the individual's capacity. Second, the lists of eight short and four long sentences differ in the amount of coherence between clauses but not in the list length. There should be no difference between trial types in the usefulness of a phonological-length-based retention mechanism, but the four long sentences condition should produce higher performance if the two clauses within a long sentence are sometimes combined to form a single long chunk. We can determine whether all age groups benefit similarly from this extra structure in the four long sentences condition. Third, the lists of four short sentences and four random pseudo-sentences also are similar in phonological length but differ in coherence inasmuch as the syntactic and semantic structure that combines words into clause units is absent from the random pseudo-sentences. We can determine whether all age groups benefit similarly from this extra structure in lists of coherent sentences.

The method serves as an opportunity for evidence convergent with what Pascual-Leone (1970) obtained, using a very different method to determine the units or chunks in memory. Pascual-Leone assumed that each stimulus cue–response association taught to a child served as a unit. In contrast, we used preexisting linguistic knowledge (cf. Burtis, 1982; Tulving & Patkau, 1962) in a way that allowed us to verify the coherence of a chunk by examining the mean amount of completion of a clause. If the completion remains the same across age groups, then the difference in performance can be attributed to a difference in the number of units accessed, that is, chunks in working memory.

Table 2
Comparison of sentence conditions in terms of number of clauses, number of sentences, and overall list length.

Condition	Number of clauses	Number of sentences	List length
4 short	4	4	Short
8 short	8	8	Long
4 long	8	4	Long
4 random	(4 pseudo)	(4 pseudo)	Short

Method

Participants

Participants were 25 children in first grade (mean age = 7.73 years, *SD* = 0.21), 26 children in sixth grade (mean age = 12.43 years, *SD* = 0.39), and 24 adults (mean age = 18.37 years, *SD* = 0.49). Children were recruited from local public schools, and adult data were obtained from Gilchrist and colleagues (2008). All participants reported normal or corrected-to-normal vision and hearing. Children received $10 and a book as compensation, and adults received course credit.

Design

Four different sentence conditions were presented to participants, with each condition composed of one of three possible sentence types: short sentences, long sentences, or random pseudo-sentences. Short sentences were simple one-clause sentences (e.g., *Thieves took the painting*). Long sentences were composed of two short sentences that were meaningfully conjoined (e.g., *Our neighbor sells vegetables but he also makes fruit juice*). Finally, random pseudo-sentences were composed of words presented with little syntactic structure (e.g., *a close football your cheese*). Random pseudo-sentences were equivalent to short sentences in terms of length and were presented using a sentence intonation.

The sentence conditions were formed by varying the types of sentences presented as well as their number in some cases. A given list of sentences to be remembered could contain (a) four short sentences, (b) four long sentences, (c) eight short sentences, or (d) four random pseudo-sentences. Two trials were presented for each sentence condition, and the order of trials was randomized across participants.

Materials

The stimuli were the same as those used by Gilchrist and colleagues (2008). Spoken sentences were presented in a female voice between 45 and 70 dB. Content words within these sentences had age of acquisition norms with ratings between 100 and 350 that were found using the MRC Psycholinguistic Database (Wilson, 1988). Short sentences and random pseudo-sentences ranged from 3 to 5 words in length, and long sentences ranged from 8 to 11 words in length.

Procedure

Participants were tested one at a time, in a sound-attenuated booth equipped with a computer, headphones to listen to presented stimuli, and a microphone to record participant recall. Participants were instructed to listen carefully to the presented sentences and to recall verbally what was just heard in any order when cued to do so. Trials began with the word “Ready” on the computer screen for 1000 ms. Spoken stimuli were then presented via headphones.

To ensure that any possible differences across age groups were not due to the selection of materials, short and long sentences were reassigned to different conditions across participants as a means of counterbalancing. Long sentences presented to half of the participants were broken into two short sentences presented in different conditions to the other half (Table 1). For example, participants in the first group might be presented with the long sentence *I upset my mother when I lied about the money*. Participants in the second group would be presented with the short sentences *I upset my mother* and *I lied about the money,* with each sentence placed into different trials of a given condition (here either the four or eight short sentences condition) to reduce any potential sentence-specific effects. In a similar manner, long sentences presented to the second group were deconstructed into short sentences that were presented to the first group of participants.

The four or eight sentences in a given condition were separated by 1000-ms pauses. After the last sentence in a given condition was presented and was followed by a final 1000-ms pause, a 500-ms, 400-Hz tone was presented to cue recall. Participants then provided responses by speaking into the microphone. In each condition, participants were given 1 min to recall as many of the words as possible in any order, but they were free to terminate recall earlier via a key press. Another key press signaled when participants were ready to move to the next list of sentences.

All verbal responses were converted into sound files and saved for later transcription by the experimenter.

Analyses

We used three different measures to examine how different aspects of immediate memory for sentences changed with development: words recalled per trial, clause access, and clause completion. *Words recalled per trial* was measured as the total number of words recalled from each sentence condition. Because more than one instance of a word could be present within a trial (e.g., *the* often appeared more than once), each occurrence of that word in recall was scored as a separate word recalled.

Two special measures were used also by Gilchrist and colleagues (2008), adapted from Naveh-Benjamin, Cowan, Kilb, and Chen (2007). *Clause access* was taken as a measure of the number of independent groups presented on a trial that were retrieved successfully. We measured this as the number of clauses from which at least one content word was retrieved. This measure was based on the assumption that words recalled from one particular clause or one-clause sentence could be presumed to form part of the same chunk present within working memory. Previous research suggests that words recalled from a given clause are not typically recalled in isolation; instead, access to a clause typically entails the recall of multiple words from the clause (Gilchrist et al., 2008).

Designating the unit of analysis as a clause meant that short sentences had one such unit and long sentences had two such units. It is possible that young children are less able than older participants to join two short sentences into an overarching long sentence, in which case there would be more separate units to be recalled in young children, producing especially poor clause access in the four long sentences condition. For the random condition, each random pseudo-sentence was counted as a clause, allowing this condition to serve as a control to observe the effect of linguistic coherence in the other conditions.

Clause completion was defined as the proportion of words recalled from a clause, contingent on at least one word having been recalled. This measure was assumed to reflect the amount of interword association between words in a clause, that is, the coherence of the clause as a chunk in memory.

Results

Words recalled yields a holistic view of developmental growth in working memory, and the access and completion measures yield a more analytic understanding of that developmental growth.

Words recalled per trial

Developmental increases in mean words recalled per trial can be observed in Fig. 1. A repeated measures analysis of variance (ANOVA) of this variable included age as a between-participants factor and sentence condition as a within-participants factor. There was a significant effect of age, $F(2, 72) = 12.26$, $p < .0001$, $\eta_p^2 = .254$, with fewer words recalled in younger children ($M = 7.52$, $SD = 2.90$) than in older children ($M = 11.22$, $SD = 2.90$) or adults ($M = 10.81$, $SD = 2.48$). Post hoc Newman–Keuls tests showed significant differences between children in first grade and the two older age groups, who did not differ from each other. There was also a significant effect of condition, $F(3, 216) = 135.95$, $p < .0001$, $\eta_p^2 = .654$, with the greatest number of words recalled for four long sentences ($M = 14.30$, $SD = 5.82$), followed by eight short sentences ($M = 10.51$, $SD = 3.78$), four short sentences ($M = 9.66$, $SD = 2.60$), and four random pseudo-sentences ($M = 4.93$, $SD = 1.82$). Performance levels on all conditions were significantly different from all others in the post hoc tests except that performance on four short sentences and eight short sentences did not differ. The overall pattern of performance replicates what is found in a separate analysis of the young adults: four long > eight short = four short > four random.

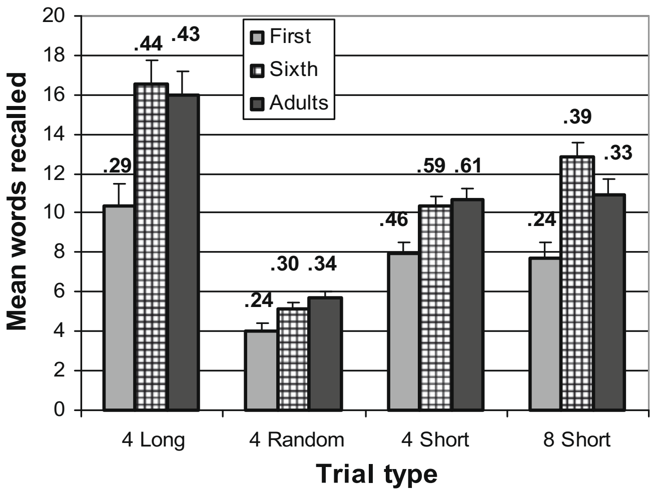

Fig. 1. Mean number of words recalled per trial for each age group. Error bars are standard errors of the mean. The number of words recalled was always well below the ceiling level. Proportions of words recalled per condition are located directly above standard error bars. First, first-graders; Sixth, sixth-graders.

The main effects were qualified by a significant Age Group Condition interaction, $F(6, 216) = 5.06$, $p < .0001$, $\eta_p^2 = .123$. Newman–Keuls tests indicated that the first-grade pattern was the same as the adult pattern (above). Interestingly, the pattern in the intermediate sixth-grade group was more differentiated: four long > eight short > four short > four random. The discrepancy occurred because the sixth-graders did better than the adults on eight short unrelated sentences. The reason for this counterintuitive difference is unclear, but it can be said, at least, that performance on our task is adult-like by sixth grade.

In addition, we examined whether qualitative patterns in recall differed across age groups in three ways. First, certain items within a sentence may have been more likely to be recalled over others. Specifically, we were interested in comparing the recall of key nouns or verbs within each clause to items that were less critical for overall comprehension (e.g., articles, prepositions). Indeed, critical items within clauses were more likely to be recalled over noncritical items, $F(1, 72) = 183.84$, $p < .001$, $\eta_p^2 = .72$ (critical, $M = .41$, $SD = .11$; noncritical, $M = .34$, $SD = .11$). Although all age groups showed a recall benefit for critical items, a significant Age $\times$ Item Type interaction, $F(2, 72) = 6.30$, $p < .005$, $\eta_p^2 = .15$, revealed that this advantage was smaller for young children relative to the older age groups, who did not differ from each other. Means for critical and noncritical items were .32 and .28 for first-graders, .47 and .38 for sixth-graders, and .45 and .37 for adults, respectively ($SD = .11$ in each case). Thus, items are not equiprobable in their likelihood of being recalled, and young children rely somewhat less than older participants on the items that are critical for general comprehension and gist.

Second, we investigated any instances of synonyms for the correct words recalled for each trial. In the prior analysis, synonyms were not included among words recalled. If young children were more likely than older participants to recall synonyms of words instead of the correct words, our scoring could be viewed as underestimating capacity in young children and, therefore, overestimating the age difference in capacity. Overall, instances of recalled synonyms were extremely rare, with mean recall of synonyms per trial being less than 1. There was an effect of age in terms of mean synonyms recalled, $F(2, 72) = 3.53$, $p < .05$, $\eta_p^2 = .09$, and Newman–Keuls tests indicated that more synonyms were recalled by adults ($M = .22$, $SD = .15$) than by first-graders ($M = .14$, $SD = .15$) or sixth-graders ($M = .12$, $SD = .15$), who did not differ from each other. Given that synonym recall increased with age, this finding amplifies, rather than nullifies, the age difference in words recalled verbatim.

The greatest number of synonyms recalled in place of presented words occurred for four long sentences ($M = .31$, $SD = .38$), with increasingly fewer recalled for eight short sentences ($M = .21$, $SD = .31$), four short sentences ($M = .12$, $SD = .24$), and four random pseudo-sentences ($M = .006$, $SD = .06$). This difference was statistically significant, $F(3, 216) = 17.10$, $p < .0001$, $\eta_p^2 = .191$. Post hoc Newman–Keuls tests indicated that all conditions were significantly different from each other. The interaction

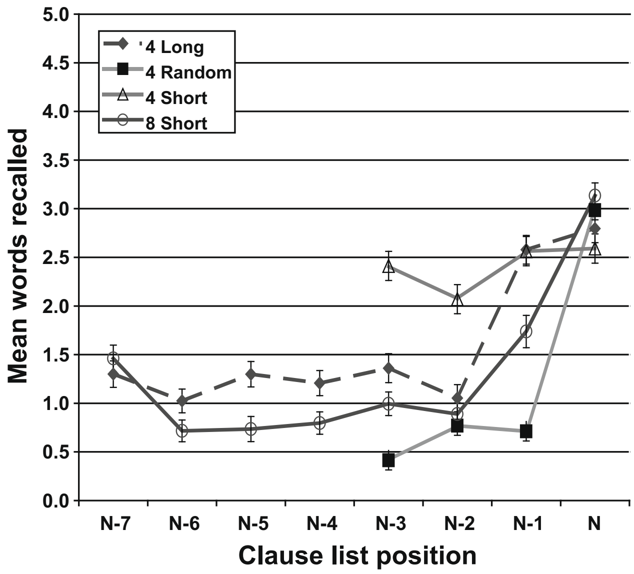

Fig. 2. Mean number of words recalled in each sentence condition by clause serial position in the presented list. Error bars are standard errors of the mean. *N* refers to the number of clauses present within a sentence list.

between age and sentence condition did not approach significance. Thus, the conditions and groups producing the most verbatim words also produced the most synonyms.

Finally, it was possible that different patterns of recall across serial positions could have been present in the different age groups. For this reason, we examined number of words recalled per clause serial position for each age group and sentence condition. It was assumed that if young children differed from older age groups in serial position effects, this would be manifest as an Age × Clause Position interaction for each sentence condition. However, for all conditions, repeated measures ANOVAs produced no significant interactions of Age × Clause Position: four long, $F(14, 504) = 1.18$, $p = .29$; four random, $F(6, 216) < 1$, $p = .56$; four short, $F(6, 216) < 1$, $p = .71$; eight short, $F(14, 504) = 1.27$, $p = .22$. In general, regardless of age group, words contained in list-final clauses were most likely to be recalled (see Fig. 2), with words recalled from four short sentences as the sole exception: four long, $F(7, 504) = 32.12$, $p < .0001$, $\eta_p^2 = .308$; four random, $F(3, 216) = 118.47$, $p < .0001$, $\eta_p^2 = .622$; four short, $F(3, 216) = 2.21$, $p = .09$; eight short, $F(7, 504) = 43.38$, $p < .0001$, $\eta_p^2 = .376$. This suggests that the significant age differences that we found were not influenced by changes in serial position effects with age.

Clauses accessed per trial

As in the previous study, we operationally defined the chunk as the most coherent unit presented, the clause or short sentence. In the case of random pseudo-sentences, we defined a chunk as a presented pseudo-sentence, equivalent in length to a one-clause short sentence. Fig. 3 shows that there were age differences in the number of clauses accessed. The ANOVA showed a significant effect of age group, $F(2, 72) = 13.08$, $p < .0001$, $\eta_p^2 = .267$. Newman–Keuls tests indicated that first-graders accessed fewer clauses ($M = 2.34$, $SD = 0.79$) than did sixth-graders ($M = 3.37$, $SD = 0.79$) or adults ($M = 3.28$, $SD = 0.79$), who did not differ from each other.

There was also a significant effect of sentence condition, $F(3, 216) = 117.08$, $p < .0001$, $\eta_p^2 = .619$, with the greatest number of clauses accessed from four long sentences ($M = 4.16$, $SD = 1.49$), followed by eight short sentences ($M = 3.19$, $SD = 1.12$), four short sentences ($M = 2.83$, $SD = 0.67$), and four random pseudo-sentences ($M = 1.80$, $SD = 0.65$). The post hoc tests showed that all conditions were significantly different from each other.

The above effects were qualified by a significant interaction, $F(3, 216) = 5.28$, $p < .0001$, $\eta_p^2 = .128$. Newman–Keuls tests indicated an interesting pattern. In the two older groups, the pattern was the same as for words recalled: adults, four long > eight short = four short > four random; sixth-graders, four long > eight short > four short > four random. In first-graders, however, there was little difference

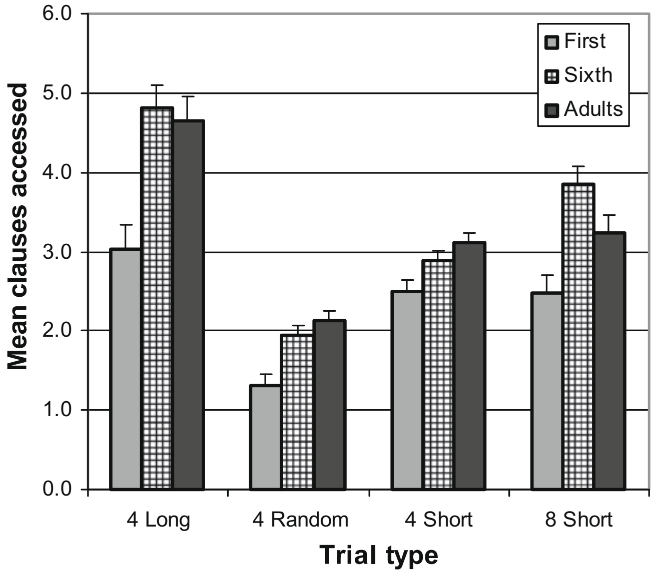

Fig. 3. Mean number of clauses accessed by condition for each age group. Error bars are standard errors of the mean. First, first-graders; Sixth, sixth-graders.

between sentence types: four long = eight short = four short > four random. The absence of a difference between four long sentences and eight short sentences indicates that first-graders did not benefit significantly in clauses accessed from the association between clauses in the four long sentences condition.

Clause completion

As mentioned above, we defined clause completion as the proportion of words recalled from a given clause given that it had already been accessed (i.e., conditional on at least one content word having been recalled). Importantly, as Fig. 4 illustrates, we found no significant effect of age on clause completion, $F(2, 70) = 1.22$, $p = .30$, $\eta_p^2 = .034$. Power analyses using G*Power software (Faul, Erdfelder, Lang, & Buchner, 2007) showed that we were able to detect an effect of $\eta_p^2 = .08$, a small effect, with a power of .80.

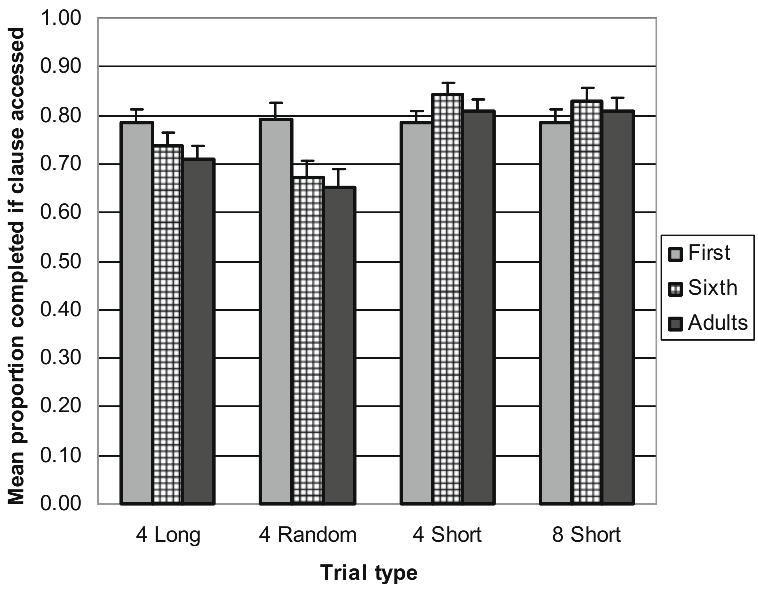

Fig. 4. Mean proportion of words recalled conditional on clauses that were accessed for each age group. Error bars are standard errors of the mean. First, first-graders; Sixth, sixth-graders.

There was a significant effect of condition, $F(3, 210) = 12.97$, $p < .0001$, $\eta_p^2 = .156$. Completion rates were highest for four short sentences ($M = .81$, $SD = .12$) and eight short sentences ($M = .81$, $SD = .13$), followed by four long sentences ($M = .74$, $SD = .13$) and four random pseudo-sentences ($M = .71$, $SD = .18$). Shorter sentences, by virtue of their simplicity as well as their brevity, are easier to complete than long sentences that contain additional words or random pseudo-sentences that are incoherent in their meaning and structure. Newman–Keuls tests showed only that the lists of shorter sentences (either four or eight short sentences) had higher levels of completion than did the lists of either long or random incoherent sentences.

It might seem to be remarkable that four random pseudo-sentences could be completed about as well as four long sentences. The way this occurred is that it was usually the most recent pseudo-sentence from which most words were recalled, presumably making use of phonological memory. The equivalence of four short and eight short sentences at a high rate of completion (.81) strengthens the assumption that these short sentences function as integrated units or chunks in working memory.

This effect was qualified by a significant Age × Condition interaction, $F(6, 210) = 3.50$, $p < .01$, $\eta_p^2 = .091$, in contrast to Gilchrist and colleagues (2008), who found no significant differences in patterns of completion between young and older adults. Newman–Keuls analyses showed that, to a first approximation, both young adults and sixth-graders conformed to the pattern of four short = eight short > four long = four random. (The groups differed slightly in that, in adults, the four long vs. four short comparison did not reach significance; whereas in sixth-graders, the four long vs. eight short comparison did not reach significance.) In first-graders, in contrast, there was no difference between conditions (see Fig. 4). Linguistic structure did not aid the completion rate in first-graders.

Discussion

We examined whether the increase in working memory performance accompanying development was driven by an increased ability to retain more chunks or to form larger chunks. The number of chunks was estimated as the number of unrelated short one-clause sentences at least partly recalled (i.e., accessed) from lists of such sentences. The size of each chunk was estimated as the proportion of words recalled from an accessed clause. The results indicated a developmental increase in the number of independent chunks recalled. In contrast, chunk completion, the proportion of words recalled for a given accessed clause, remained at a similar level across the age groups tested. In our further discussion, we first examine the effects of our manipulations of the stimulus type on three dependent measures so as to understand what they can tell us about the memory processes for our materials. Then we examine the basis of developmental changes that were obtained.

Effects of manipulations

Effects on words recalled

It is clear that increasing the amount of information present within a given word sequence facilitates recall (Fig. 1). Across all age groups, when sequences contained no linguistic coherence, as in the four random pseudo-sentences condition, the number of words recalled was small and roughly corresponded to the limits of working memory capacity to be expected given previous work (Baddeley, 1986; Broadbent, 1975; Cowan, 2001; Cowan et al., 2005; Miller, 1956) under the assumption that each word functioned as a separate chunk. Specifically, the number recalled was roughly 4 in young children and increased to roughly 6 in adults, approximating the usual limit if phonological memory for sequences can make a contribution.

With the linguistic coherence provided by sentence structure in the other conditions, the numbers of words recalled per trial were increased substantially, as one would expect under the assumption that each chunk unit in working memory then includes multiple words such as the words within a clause. Linguistic structure in the short sentences roughly doubled the number of words recalled compared with random pseudo-sentences. The reason why it did not increase the number of words recalled even more may be that the phonological memory trace makes little contribution for stimulus materials exceeding a spoken duration of approximately 2 s (Baddeley, 1986).

Another indication of the benefit of linguistic coherence across all age groups is that the number of words recalled was highest when the words were presented within long sentences. The eight short sentences condition contained the same number of clauses as the four long sentences condition but did not provide a similar advantage in recall. This difference between conditions shows that connecting information meaningfully across clauses helps to increase the number of words that can be remembered. However, the facilitation of recall that a long sentence provided is less than one would expect if each long sentence were encoded as a single chunk. In particular, although the four long sentences condition contains double the number of words present within the four short sentences condition, there was less than a two-to-one ratio of words recalled in these two conditions for every age group. It appears that the association between clauses was sometimes, but not always, of use, a conclusion that is further supported below.

Effects on clause access

Below, clause completion provides evidence for the clause as a functional unit or chunk. Assuming that to be the case for the time being, results from the proportion of clauses accessed (i.e., clauses from which at least one substantive word was recalled) can be taken as evidence for a constant capacity in working memory. Between 2 and 4 clauses were accessed in the four and eight short sentences conditions (Fig. 2), as would be expected according to previous developmental evidence on capacity limits (e.g., Cowan, 2001; Cowan et al., 2005). The finding that the number of clauses accessed was independent of the number of sentences is an important confirmation that a mnemonic process limited by the number of chunks is at work here. The long length of the sentence lists puts them out of range for there to be much of a contribution of phonological memory, and the resulting number of clauses accessed is similar to the core verbal working memory capacity that can be obtained by requiring articulatory suppression during recall (Chen & Cowan, 2009).

In the case of random pseudo-sentences, individual words may function as separate chunks, so the method of scoring each pseudo-sentence as a single clause resulted in very few (1–2) such pseudo-sentences being accessed. Therefore, this condition effectively served as a control to observe the effect of removing syntactic and semantic associations between items in a clause.

As was found in the number of words recalled, access to clauses was facilitated by additional semantic and syntactic structure. Thus, access for clauses within four long sentences was greater than that within eight short sentences despite both conditions containing the same number of independent clauses. We believe that, in this case, pairs of clauses within a long sentence sometimes function as a single chunk.

Effects on clause completion

Results from clause completion suggested that a one-clause sentence generally did serve as a functional chunk in working memory. In the four and eight short sentences conditions, in which each sentence had one clause, we found high levels of clause completion (~80%) for all age groups (Fig. 4) despite proportions of clause access well below 80%. This suggests strong associations between words within a sentence and much weaker associations between words in different unrelated sentences, a pattern that validates the idea that the short sentence generally serves as a single chunk in this experiment (cf. Cowan, 2001).

In the older groups, the level of clause completion was somewhat lower for the remaining conditions, but this age-related trend is discussed below in the section on developmental effects. Notice that the overall pattern is one in which the number of units accessed varies widely across conditions, whereas the completion of accessed units varies only slightly across conditions. This organization of recall into large sequences could be driven at least in part by syntactic and semantic coherence in the linguistic conditions versus only phonological memory and prosodic coherence in the four random pseudo-sentences condition.

Developmental effects

Previous research suggests that, relative to adults, children have reduced working memory capacity as well as a reduced immediate memory span (e.g., Cowan et al., 2005). As mentioned above, these developmental differences may be due either to young children having a smaller number of chunks that can be held in memory or to them forming smaller chunks. For the lists of simple linguistic materials that we used, the results provide support for the former factor (number of chunks retained) over the latter factor. The meaning of the developmental change in each of three dependent measures is discussed in turn.

Development of words recalled

We found evidence for a developmental increase in the number of words recalled in all sentence conditions (Fig. 1), consistent with the well-known increase in span with development. Older children and adults generally recalled more words per condition than did young children.

Young children were able to take advantage of semantic or syntactic structure. Like the older groups, young children recalled approximately twice as many words from short sentences as they did from random pseudo-sentences. Structure provided by long sentences also was used by all age groups, who recalled more words from four long sentences than from either short sentences condition. The advantage for long sentences, however, was somewhat smaller in young children than in the older groups, suggesting that young children were somewhat less likely than older participants to combine the clauses of a long sentence into a single chunk.

In general, developmental studies of memory that have documented an increase in span with age have typically used word recall as a measure of interest. Although our results provide additional confirmation of these prior findings, examining recalled words provides insufficient understanding of potential mechanisms that underlie the characteristic increase in memory span. For this reason, we also examined developmental differences in clause access and completion. These measures can help to determine whether any age differences observed reflect changes in the number or size of chunks maintained in working memory.

Development of clause access

We found a developmental increase in the number of independent units that could be retrieved. With access to a clause taken as memory of one chunk (even though it was often an imperfect memory of the presented chunk), the number of remembered chunks increased with age (Fig. 2). The increase in the number of recalled chunks from approximately 2.5 in young children to 3 or 4 in older children and adults accords well with previous observations (e.g., Chen & Cowan, 2009; Cowan et al., 2005).

The increase in clause access with age was not dependent on a certain level of structure; it occurred in all four conditions. Nevertheless, there is evidence that the use of information from longer structures increased with age. The increase in the number of clauses accessed in long sentences compared with short sentences was significant by post hoc tests in the two older age groups but not in the young children. This is perhaps further evidence that the appreciation of subtle details of linguistic structure continue to develop during the elementary school years (e.g., Chomsky, 1969).

Clause completion

Clause completion in older children and adults changed across sentence conditions (Fig. 4), implying sensitivity to the linguistic structure. Compared with short sentences, there was a diminished completion of "clauses" in the form of random pseudo-sentences, as one might expect if more than one slot in working memory needed to be used to retain the words that were recalled from a single pseudo-sentence. There was also a diminished completion of clauses from long sentences. Perhaps these older participants attempted to use the associations between clauses within a long sentence to retain more clauses, which could occur at the expense of some within-clause information.

Interestingly, as shown in Fig. 4, younger children showed no sensitivity to linguistic structure in their clause completion. As noted previously, there were no age differences in serial position effects that might account for this difference in the use of linguistic structure. It is possible that the intonation structure was more important than high-level information for chunk formation in young children (cf. Cowan, 1989). Their focusing on intonation structure typically led them to recall most of the words from a single pseudo-sentence, whereas older participants were more likely to sacrifice words from a pseudo-sentence so as to recall something from a larger number of pseudo-sentences or to sacrifice words from clauses within a long sentence so as to recall something from a larger number of such clauses. Given that the sentences were well within young children's ability, the absence of an effect of sentence type on clause completion is an interesting and new finding.

Despite these somewhat subtle age differences in the effects of conditions, there was no age difference in the mean level of clause completion across conditions. This result extends the findings of Gilchrist and colleagues (2008), who obtained no effect of adult aging on clause completion, only on clause access.

Conclusions

As children get older, the amount of information that can be held in memory increases. This finding is robust and has been observed in many developmental studies. Yet it has often been unclear exactly what factors underlie developmental differences in memory span. Using measures not only to examine words recalled in spoken sentences but also to examine how many independent one-sentence clauses came to mind (clause access) and how much of each such clause was recalled (clause completion), we found large developmental differences in the level of clause access but not in clause completion. From these results, we propose that an increase in the number of chunks that can be stored in working memory underlies developmental differences in immediate recall, converging with the developmental increase proposed by several others (e.g., Burtis, 1982; Cowan et al., 2005; Johnson, Im-Bolter, & Pascual-Leone, 2003; Pascual-Leone, 1970, 2005).

We emphasize the importance of these findings because the results converge with findings from previous studies despite very different experimental methodologies and assumptions. Studies by neo-Piagetians showing changes with age in M-space often relied on a priori assumptions of how chunks were formed based on the number of units that were presented in the stimuli (e.g., Burtis, 1982; Pascual-Leone, 1970). In contrast, our clause completion measure allows us to verify that the mean size of the units in memory did not differ by age group. Despite these methodological differences, our clause access measure confirms the conclusion that there are developmental improvements in immediate memory coming from an increase in the number of units that can be maintained. Contrary to some other results (e.g., Ottem et al., 2007), we found no evidence for an increase in chunk size with age. This outcome might, of course, be different for more complex materials.

Acknowledgments

This research was conducted with support from National Institutes of Health (NIH) Grant R01-HD21338. We thank Angela AuBuchon, Melissa Knipe, Caleb O'Brien, and Christopher Zwilling for assistance and J. Scott Saults for programming.

References

Allen, R. J., & Baddeley, A. D. (2009). Working memory and sentence recall. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 63–85). Hove, UK: Psychology Press.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575–589.

Broadbent, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory* (pp. 3–18). New York: John Wiley.

Burtis, P. J. (1982). Capacity increase and chunking in the development of short-term memory. *Journal of Experimental Child Psychology, 34*, 387–413.

Caplan, D., Waters, G., & DeDe, G. (2007). Specialized verbal working memory for language comprehension. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (pp. 272–302). New York: Oxford University Press.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386–404.

Chen, Z., & Cowan, N. (2005). Chunk limits and length limits in immediate recall: A reconciliation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1235–1249.

Chen, Z., & Cowan, N. (2009). Core verbal working memory capacity: The limit in words retained without covert articulation. *Quarterly Journal of Experimental Psychology, 62*, 1420–1429.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, NJ: Lawrence Erlbaum.

Chomsky, C. S. (1969). *The acquisition of syntax in children from 5 to 10*. Cambridge, MA: MIT Press.
Cowan, N. (1989). Acquisition of Pig Latin: A case study. *Journal of Child Language, 16*, 365–386.
Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.
Cowan, N., & Alloway, T. (2009). The development of working memory. In M. Courage & N. Cowan (Eds.), *The development of memory in infancy and childhood* (pp. 303–342). Hove, UK: Psychology Press.
Cowan, N., Chen, Z., & Rouder, J. (2004). Constant capacity in an immediate serial recall task: A logical sequence to Miller (1956). *Psychological Science, 15*, 634–640.
Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., et al (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*, 42–100.
Cowan, N., Naveh-Benjamin, M., Kilb, A., & Saults, J. S. (2006). Life-span development of visual working memory: When is feature binding difficult? *Developmental Psychology, 42*, 1089–1102.
Cowan, N., Nugent, L. D., Elliott, E. M., Ponomarev, I., & Saults, J. S. (1999). The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development, 70*, 1082–1097.
Daneman, M., & Case, R. (1981). Syntactic form, semantic complexity and short-term memory: Influences on children's acquisition of new linguistic structures. *Developmental Psychology, 17*, 367–378.
Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89*, 63–100.
Faul, F., Erdfelder, E., Lang, A-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.
Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development, 37*, 283–299.
Gilchrist, A. L., Cowan, N., & Naveh-Benjamin, M. (2008). Working memory capacity for spoken sentences decreases with adult ageing: Recall of fewer but not smaller chunks in older adults. *Memory, 16*, 773–787.
Glanzer, M., & Razel, M. (1974). The size of the unit in short-term storage. *Journal of Verbal Learning and Verbal Behavior, 13*, 114–131.
Harris, G. J., & Burke, D. (1972). The effects of grouping on short-term recall of digits by children: Developmental trends. *Child Development, 43*, 710–716.
James, W. (1890). *The principles of psychology*. New York: Henry Holt.
Johnson, J., Im-Bolter, N., & Pascual-Leone, J. (2003). Development of mental attention in gifted and mainstream children: The role of mental capacity, inhibition, and speed of processing. *Child Development, 74*, 1594–1614.
Kail, R. (1992). Processing speed, speech rate, and memory. *Developmental Psychology, 28*, 899–904.
Kail, R., & Hall, L. K. (2001). Distinguishing short-term memory from working memory. *Memory & Cognition, 29*, 1–9.
Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.
Naveh-Benjamin, M., Cowan, N., Kilb, A., & Chen, Z. (2007). Age-related differences in immediate serial recall: Dissociating chunk formation and capacity. *Memory & Cognition, 35*, 724–737.
Ornstein, P. A., Naus, M. J., & Liberty, C. (1975). Rehearsal and organizational processes in children's memory. *Child Development, 46*, 818–830.
Ottem, E. J., Lian, A., & Karlsen, P. J. (2007). Reasons for the growth of traditional memory span across age. *European Journal of Cognitive Psychology, 19*, 233–270.
Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica, 32*, 301–345.
Pascual-Leone, J. (2005). A dialectical constructivist view of developmental intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 177–201). Thousand Oaks, CA: Sage.
Simon, H. A. (1974). How big is a chunk? *Science, 183*, 482–488.
Towse, J. N., Hitch, G. J., & Skeates, S. (1999). Developmental sensitivity to temporal grouping effects in short-term memory. *International Journal of Behavioral Development, 23*, 391–411.
Tulving, E., & Patkau, J. E. (1962). Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canadian Journal of Psychology, 16*, 83–95.
Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine readable dictionary, Version 2. Behavioural Research Methods, Instruments, and Computers, 20, 6–11. Available from http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm.