Relation of working memory to off-line and real-time sentence processing in children with specific language impairment

JAMES W. MONTGOMERY
University of North Carolina, Chapel Hill

ADDRESS FOR CORRESPONDENCE
James W. Montgomery, Speech and Hearing Sciences, University of North Carolina, Chapel Hill, NC 27599-7255. Email: jmontgomery@css.unc.edu

ABSTRACT
In this study we examined the influence of working memory on the off-line and real-time sentence comprehension/processing of children with specific language impairment (SLI). A total of 12 children with SLI, 12 normally developing children matched for chronological age (CA), and 12 children matched for receptive syntax (RS) completed three tasks. In the working memory task, children recalled as many words as possible under three processing load conditions varying in the number of mental operations (i.e., no load, single load, dual load). In the off-line comprehension task, children listened to linguistically nonredundant and redundant sentences. In the real-time sentence processing task, children monitored sentences for the occurrence of a target word appearing at the beginning, middle, or end of a test sentence and pushed a response pad as quickly as possible upon hearing the target. In the memory task, SLI children recalled fewer words in the dual-load condition relative to CA peers, who showed no condition effect. The SLI and RS groups performed similarly overall; however, both groups recalled fewer words in the dual-load condition than in the other conditions. In the off-line task, the SLI group comprehended fewer sentences of both types relative to the CA controls and fewer redundant sentences relative to themselves and to the RS controls. A significant correlation between working memory and sentence comprehension was found for the SLI group and control groups. For the on-line task, between-group analyses revealed that the SLI group yielded an overall slower word recognition reaction time than the CA and RS groups. Working memory and sentence processing were not correlated for any group. Results were interpreted to suggest that SLI children have a more limited functional working memory capacity than their CA peers. Children with SLI also appear to have greater difficulty managing their working memory resources relative to both age peers and younger children when performing a conventional off-line sentence comprehension task but not a real-time sentence processing task.

It has been proposed that the language impairment of many children with specific language impairment (SLI) may be related to some extent to some kind of memory deficiency (e.g., Ceci, Ringstrom, & Lea, 1981; Curtiss & Tallal, 1991; Graham, 1980; Montgomery, 1995a, in press). Ample experimental evidence reveals that many SLI children demonstrate deficits in a number of major func-
tions of verbal short term memory, including scanning speed (Sinninger, Klatzky, & Kirchner, 1989), retrieval (Ceci et al., 1981; Gillam, Cowan, & Day, 1995; Gillam, Cowan, & Marler, 1998; Kail, Hale, Leonard, & Nippold, 1984), and capacity (Gathercole & Baddeley, 1990a; Kirchner & Klatzky, 1985; Montgomery, 1995b, in press). However, few studies have explored the nature of the relation between deficient verbal short term memory and language abilities in these children – and then only with regard to their off-line sentence comprehension abilities (e.g., Montgomery, 1995a, in press). The potential relation between working memory and real-time (on-line) language processing has not been investigated. Because much of language understanding occurs in the moment, it would appear that such an investigation would be important to shed light on the language comprehension abilities and difficulties of these children. Accordingly, in the present study we examined whether potential differences in the functional working memory capacity (e.g., Daneman & Carpenter, 1983; Just & Carpenter, 1992) of children with and without SLI were associated with both off-line and on-line sentence comprehension.

The two most prominent theories of working memory include Baddeley’s (1986) phonological loop model and the model of Daneman and Carpenter and their associates (Daneman & Carpenter, 1983; Daneman & Merikle, 1996; Just & Carpenter, 1992; King & Just, 1991). According to Baddeley, working memory is a multicomponent, resource-limited system comprising a controlling “central executive” which is subdivided into two distinct “slave” systems: the articulatory loop system and the visuospatial scratch pad. The central executive, the least understood component of the model, is thought to regulate information flow within working memory, the retrieval of information from other memory systems, and the processing and storage of information. The articulatory loop is assumed to comprise a capacity-limited phonological short-term store and an articulatory control process (i.e., verbal rehearsal) that acts to refresh and maintain speech material in the store for a brief period. The articulatory loop is thought to be responsible for the temporary storage of verbal information while other cognitive tasks, such as verbal reasoning or auditory and reading comprehension, are performed. Considerable empirical support has been offered by Baddeley and associates for the construct of the phonological loop using a variety of tasks. Moreover, he has shown an association between phonological working memory and a variety of linguistic abilities in adults (e.g., Baddeley, Vallar, & Wilson, 1987; Vallar & Baddeley, 1984) as well as in children (e.g., Adams & Gathercole, 1995; Gathercole & Baddeley, 1990b). Disruptions in phonological loop function (i.e., reduced phonological storage capacity and/or inefficient rehearsal abilities) can often lead to compromised comprehension because insufficient amounts of incoming information cannot be immediately and readily retained in the phonological store for it to be processed (Baddeley, 1986).

Daneman and associates’ model also characterizes working memory for language as a resource-limited system that includes both storage and processing functions. Their model roughly corresponds to the part of the central executive in Baddeley’s model that deals with language comprehension. The model of Daneman and associates is a computational model in which both storage and
processing functions share a finite pool of resources during comprehension. Storage is defined as the ability to retain temporarily verbal information that has been processed already, while processing is defined as those language operations/computations that generate various types of representations (e.g., lexical, grammatical, propositional) of the input. The model also assumes that the different comprehension processes (e.g., lexical, morphological, grammatical, propositional, pragmatic) operate simultaneously to compute partial or full representations (e.g., words, grammatical structures, propositions) of the input. In the event that the resources available to the working memory system are exceeded by the storage and/or processing demands of a comprehension task, a trade-off between storage and processing will occur. For instance, some of the resources that are allocated to maintaining old representations may be reallocated to the comprehension processes, thus leading to “forgetting” of some or all of the previously processed information residing in storage.

According to Daneman and her colleagues (Daneman & Carpenter, 1983), differences among nondisordered individuals in comprehension ability reflect individual differences in the ability to coordinate the simultaneous functions of processing and storage. They have proposed that those individuals demonstrating poor comprehension typically allocate a majority of their resources to comprehension processes, leaving fewer resources for storage. Hence, these individuals have a functionally smaller temporary storage capacity. Therefore, by the time a listener reaches the end of a sentence, the representation (typically a thematic/semantic one) that he or she constructed earlier in a sentence may be forgotten (i.e., the representation no longer receives sufficient activation to remain in an active state).

Using tasks that assess the joint functions of processing and storage, as opposed to tasks that assess just storage, has been advocated by Daneman and Carpenter (1980) as the best way to determine an individual’s functional working memory capacity. To this end, they developed the working memory span task. In a typical working memory span task, subjects are asked to read or listen to sets of sentences and then after each set to recall as many sentence-final words as possible. The processing component of the task is reflected in subjects having to process the truth value of each sentence. The storage component is reflected by their having to recall as many sentence-final words as possible. Working memory is defined as the maximum number of sentences the subject can comprehend while maintaining perfect recall of the final words.

Relative to performance on traditional span measures (e.g., digit span, word span), performance on working memory tasks is much more predictive of performance on a range of language comprehension tasks. Daneman and Carpenter (1980, 1983) found that differences in working memory predicted performance on specific reading comprehension tasks in college students. Similarly, King and Just (1991) showed that individual differences among college students in working memory capacity are highly associated with comprehension of syntactically complex sentences: “low-capacity” students scored worse than “high-capacity” students. Daneman and Merikle (1996), in a meta-analysis, demonstrated the predictive value of their working memory task (and variants of it) to performance on a range of comprehension tasks. School-age children’s working
memory abilities, however, have received far less attention. Gaulin and Campbell (1994), using a working memory task similar to Daneman and Carpenter, found that functional working memory capacity in children between 6 and 12 years of age increases up through the age of 10. While these investigators found that vocabulary level was significantly correlated with performance on the working memory measure, they did not examine the relation of working memory to sentence comprehension.

Research examining the working memory abilities of SLI children has been conducted primarily within Baddeley’s (1986) theoretical framework of working memory. Results from these few studies have been interpreted by some to suggest that SLI children have a more limited phonological memory capacity than their same-age peers (Gathercole & Baddeley, 1990a) and younger, language-matched counterparts (Montgomery, 1995a). The phonological working memory capacity of SLI children has been investigated using a nonsense word repetition task (Edwards & Lahey, 1998; Gathercole & Baddeley, 1990a; Montgomery, 1995a). It has been argued that a nonsense word repetition task in which subjects repeat nonsense words varying in length is a “purer” measure of phonological working memory capacity than a task in which subjects repeat real words (Gathercole & Baddeley, 1990a, 1990b; Henry & Millar, 1991). The reason is that successful nonsense word repetition requires listeners to invoke various phonological processes (e.g., perception, encoding, storage, retrieval, and production) independent of lexical knowledge (see Dollaghan, Biber, & Campbell, 1993, for an account of prosodic influences on nonsense word repetition). Poorer repetition of longer nonsense words than shorter nonsense words presumably reflects the capacity-limited nature of the phonological store (Gathercole & Baddeley, 1990b). The results of several studies have revealed that, relative to same-age peers (Edwards & Lahey, 1998; Gathercole & Baddeley, 1990b) and language-matched peers (Montgomery, 1995a), SLI children have greater difficulty reproducing “longer” versus “shorter” nonsense words. Such results have been interpreted to suggest that these children have a limitation in phonological memory capacity (for a detailed description of a different interpretation, see Edwards & Lahey, 1998; van der Lely & Howard, 1994). Based on the results of three follow-up experiments, Gathercole and Baddeley argued that the poorer nonsense word repetition of the SLI children was not attributable to poorer perceptual processing, verbal rehearsal, or speech production. They did suggest, however, that their capacity limitation might be related to a difficulty initially forming adequate phonological representations or to faster decaying representations. In general, similar results and interpretations have been reported by Montgomery (1995a) and Edwards and Lahey (1998).

The apparent phonological working memory capacity limitation of SLI children has been shown to be associated with their poorer sentence comprehension abilities (Montgomery, 1995b). The assumption behind the Montgomery study was that sentence comprehension must to some extent rely on a listener’s ability to store verbal input long enough to process it and integrate it with previously processed information (Baddeley, 1986; Daneman & Carpenter, 1980, 1983). In his study, Montgomery examined the relation between phonological working memory capacity and sentence comprehension in a group of school-age SLI
Montgomery: Working memory and sentence processing

children and a group of younger children matched for receptive syntax. Children completed a nonsense word repetition task (i.e., index of phonological working memory capacity) and a sentence comprehension task. The sentence comprehension task included two sets of 20 sentences each, corresponding to a set of linguistically redundant (longer) sentences and a set of linguistically nonredundant (shorter) sentences. The two sentence types were nearly structurally identical and encoded essentially the same semantic information. The only difference between the sentence types was the inclusion of redundant syntactic markings or extra words in the redundant sentences. Results revealed that the SLI children comprehended fewer longer sentences than shorter sentences relative to themselves and to their language-matched peers; the control children comprehended a comparable number of long and short sentences. In addition, a positive correlation (+.62) was found between performances on the nonsense word repetition task and the sentence comprehension task. Montgomery interpreted these findings to suggest that the more limited phonological working memory capacity of the SLI children compromised their sentence comprehension efforts. Because of their capacity limitation, these children were not able to store as much information at any given moment during processing, thus hindering their ability to construct a full sentence interpretation.

In a second study, Montgomery (in press) again examined the potential relation between working memory and off-line sentence comprehension in SLI children and two groups of control children, one matched for age and the other for single-word receptive vocabulary. Unlike the first study by Montgomery (1995b), in this study Montgomery investigated the association between functional working capacity (e.g., Daneman & Carpenter, 1980, 1983) and sentence comprehension. A modified Daneman and Carpenter listening span task (see the Method section for details of the task) was used to assess children’s functional working memory capacity. In this task, children were presented with word lists comprising highly familiar, monosyllabic words. Lists varied from three to seven words. After each list, children recalled as many words as possible. Word lists were presented under three conditions designed to assess how storage was affected by variation in processing load. Under the no-load condition (simple span), children recalled as many words as possible with no regard to order of presentation. Under the single-load condition (size processing), children recalled as many words as possible according to the physical size of the word referents, beginning with the “smallest thing” and ending with the “biggest thing.” Under the dual-load condition (semantic processing + size processing), children recalled as many words as possible according to both semantic category and physical size of the word referents within each category. The children also participated in a sentence comprehension task, identical to the one used by Montgomery (1995b), in which they listened to nonredundant and redundant sentences. Results of the memory task showed that the subject groups yielded comparable levels of recall in the no-load and single-load conditions. The age controls demonstrated comparable recall across conditions. However, both the SLI and younger children recalled significantly fewer words in the dual-load condition relative to the no-load and single-load conditions. On the comprehension task, the SLI children relative to the age controls demonstrated poorer recall
of both nonredundant and redundant sentences. Relative to the younger children, the SLI children comprehended a comparable number of nonredundant sentences but significantly fewer redundant sentences. More important, whereas the age controls and younger children comprehended a comparable number of sentence types, the SLI children comprehended significantly fewer redundant than nonredundant sentences. The results were interpreted to suggest that, relative to both age peers and younger children, SLI children have greater difficulty managing their verbal working memory resources during a conventional, off-line comprehension task.

Conventional comprehension tasks are post-sentence indices of comprehension. Understanding is not assessed as it unfolds in the moment but instead after stimulus presentation. Thus, a number of “extraneous” mental operations may intervene between the input and a listener’s response, thereby obscuring identification of the psycholinguistic processes that underlie comprehension. For instance, successful performance on an off-line comprehension measure such as a conventional picture-pointing task presumably depends on the listener coordinating a number of working memory and other information processing abilities. Listeners first must comprehend and store the stimulus sentence. They then must presumably generate and store a linguistic representation corresponding to each foil picture. Then they must compare the input sentence representation with the representation of each foil picture, while simultaneously scanning the pictures. Finally, they must select the one picture that matches the input sentence. Thus viewed, poor sentence comprehension could result from either deficient linguistic knowledge or inefficient working memory/information processing.

The relation between functional working memory and real-time sentence processing has been examined in adults (e.g., Kempler, Almor, Tyler, Andersen, & MacDonald, 1998; King & Just, 1991; Waters & Caplan, 1997), but it has not been investigated in children. The focus of real-time language processing studies is to allow investigators to examine the unconscious mental representations and operations that are automatically invoked during the course of comprehension (Tyler, 1992), uncontaminated by off-line processes and strategies. In light of the association between working memory and off-line sentence comprehension observed in SLI children and their normally developing peers, one aim of the present study was to examine whether there might also be a relation between working memory and real-time sentence processing.

The relation between working memory and real-time sentence processing in SLI children is an important area of inquiry for a couple of reasons. Theoretically, results from such studies would shed light on the psycholinguistic processes (including the potential role of working memory) underlying children’s language comprehension and how all children (not just those with SLI) are able to manage their cognitive resources during the processing of spoken language. Clinically, a greater range of more sensitive assessment and intervention methods may be developed with a better understanding of the psycholinguistic processes inherent in a variety of language comprehension/processing tasks.

The few studies that have examined the real-time processing of spoken language by SLI children have shown them to process language more slowly than their age peers (Montgomery & Leonard, 1998; Stark & Montgomery, 1995)
and some younger, language-matched peers (Montgomery, Scudder, & Moore, 1990). In these word recognition reaction time studies, children monitored sentences for the occurrence of a target word and pressed a response pad as quickly as possible upon hearing the target. Based on the pattern of results from these studies, Montgomery and colleagues, from within the revised cohort theory of word recognition (Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978), have interpreted the slower language processing of SLI children as a reflection of inefficient lexical retrieval. Relative to their peers, SLI children are thought to be slower at retrieving the linguistic properties of incoming words and/or evaluating their properties relative to prior context, as opposed to having difficulty with rapidly mapping the acoustic signal onto stored lexical representations (Montgomery, 1999; Montgomery et al., 1990; Stark & Montgomery, 1995). It is not known, however, if the working memory limitations of SLI children may also play a part in their inefficient language processing.

This study had two aims. First, we wanted to see if the finding that SLI children, relative to age peers, have reduced functional working memory capacity (Montgomery, in press) would be replicated. Second, we were interested in determining whether a differential relation between working memory and real-time sentence processing and off-line sentence processing might exist.

METHOD

Participants

A total of 36 children participated in this study: 12 with SLI (mean age = 9;1), 12 with normal language matched for chronological age (CA; mean age = 8;11); and 12 with normal language matched for level of receptive syntax knowledge (RS; mean age = 7;2). The children with SLI achieved a score that was more than −1 SD from the mean on at least two of three subtests on the receptive portion and expressive portion of the Clinical Evaluation of Language Fundamentals—Revised (CELF-R) (Semel, Wiig, & Secord, 1987). They also obtained an overall receptive language score and an overall expressive language score on the CELF-R that fell more than 1 SD below the mean. They also performed greater than −1 SD from the mean on the Test of Reception of Grammar (TROG) (Bishop, 1989).

The CA and RS children performed at or above 1 SD from the mean on the same language measures. Children’s single-word receptive vocabulary knowledge was assessed using the Peabody Picture Vocabulary Test—Revised (PPVT-R) (Dunn & Dunn, 1981), although no performance criterion was set for entrance into the study. Each child with SLI was matched with a same-gender CA child ± 3 months and a same-gender RS child on the number of blocks passed on the TROG.

All children demonstrated normal range nonverbal IQ (85–120) on the Test of Nonverbal Intelligence (TONI) (Brown, Sherbenou, & Johnsen, 1990), normal-range hearing sensitivity as determined by audiometric puretone screening at 20 dBHL (American National Standards Institute, 1989), and normal or corrected vision. They also performed at or above the 67th percentile on the Gold-
Table 1. Chronological age (months); PPVT mean standard score (SS) and mean raw score (RS); mean receptive language score (RLS) and mean expressive language score (ELS) on the CELF-R; mean number of blocks passed and mean standard score on the Test of Reception of Grammar (TROG); and mental quotient (IQ) on the Test of Nonverbal Intelligence for the SLI, CA, and RS children

<table>
<thead>
<tr>
<th>Group</th>
<th>Age (mos.)</th>
<th>PPVT</th>
<th>RLS</th>
<th>ELS</th>
<th>BLKS</th>
<th>SS</th>
<th>IQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI</td>
<td>M</td>
<td>109.3</td>
<td>86.5</td>
<td>75.5</td>
<td>69.4</td>
<td>12.6</td>
<td>81.5</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>10.7</td>
<td>10.3</td>
<td>4.5</td>
<td>7.7</td>
<td>1.2</td>
<td>1.7</td>
</tr>
<tr>
<td>CA</td>
<td>M</td>
<td>107.6</td>
<td>105.5</td>
<td>108.9</td>
<td>107.3</td>
<td>17.0</td>
<td>106.3</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>9.9</td>
<td>6.8</td>
<td>9.3</td>
<td>7.9</td>
<td>1.5</td>
<td>9.0</td>
</tr>
<tr>
<td>RS</td>
<td>M</td>
<td>86.8</td>
<td>104.1</td>
<td>104.1</td>
<td>105.5</td>
<td>12.7</td>
<td>97.9</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>5.3</td>
<td>9.5</td>
<td>5.9</td>
<td>5.9</td>
<td>.81</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Man–Fristoe Test of Articulation (Goldman & Fristoe, 1986) and demonstrated no oral structural or motor impairments affecting speech or nonspeech movements of the articulators (Robbins & Klee, 1987). No child had a history of frank neurologic impairment or psychological/emotional disturbance or attention deficit disorder (from parent report). English was the primary language spoken by all of the children. Subject groups differed in nonverbal IQ, $F(2, 33) = 10.62, p < .001$; SLI children yielded a lower mean IQ than both the CA and RS children (Tukey HSD, $p < .05$). For each child, the number of years of education attained by the mother was also obtained. No significant differences emerged across groups with respect to the number of mothers who attained a college education, $\chi^2(2) = 1.92, p > .05$. There were seven boys and five girls in each of the subject groups. Within the SLI group, four children were black, Hispanic, or Asian. Two minority children were included in the CA group and two in the RS group. To adjust for potential dialectical variation of the black children, responses on the expressive subtests of the CELF-R were scored relative to the dialectic forms reported by Washington and Craig (1994). Table 1 presents group cognitive and language data.

Task 1: Working memory task

The verbal working memory task was the same one used by Montgomery (in press), which was a variant of Daneman and associates’ memory task. In the Daneman task, subjects listen to increasingly longer sets of sentences and are asked to process the truth value of each sentence; they are then asked to recall
as many sentence-final words as possible. Whereas the demands for storage clearly systematically increase, given that subjects must recall progressively more sentence-final words, similar systematic increases in processing do not necessarily appear to occur. Subjects need only to process single sentences at a time. Furthermore, the number of comprehension processes required of each sentence or set of sentences may not always vary systematically. It would appear, then, that clear and independent estimates of storage and processing might not always be yielded by this task. For this reason, the present study used a three-condition working memory task in which both storage and processing varied systematically. We reasoned that, if SLI children have less functional working memory capacity than their normal developing peers, they should perform more poorly when both the storage and processing requirements are high. Moreover, we reasoned that sentence comprehension (which should demand simultaneous lexical, morphological, syntactic, and semantic processing) should be more strongly associated with the combined functions of storage and processing than storage alone.

**Stimulus words.** Five highly familiar, monosyllabic words (e.g., Moe, Hopkins, & Rush, 1982) were selected from five basic semantic categories (animals, transportation, clothing, body parts, "plant things") for a total of 25 stimulus words. No rhyming words appeared in the 25-word set. Stimulus words are presented in Appendix 1.

**Stimulus word generation procedures.** A high-quality recording of the stimulus words was made in a sound treated booth as an adult male native speaker of American English read each word. Each recorded word was then low-pass filtered at 4.5 kHz, digitized at 10 kHz, and stored on disk. From a digital representation, each word was then interactively edited to identify the exact acoustic onset and offset to yield a set of "clean" words that were used to create each of the word lists. Word lists were created by outputting from the computer (10 kHz and low-pass filtered at 4.5 kHz) each word to tape in a prescribed random order. Each stimulus word was controlled for overall relative intensity (±3 volts). A 1 second interval separated each word in a list.

**List lengths (word lists).** Five word lists were created (i.e., 3-word lists, 4-word lists, 5-word lists, 6-word lists, 7-word lists). Each word list contained words from at least two semantic categories. Each word list also included three trials. Words within each word list and within each trial were randomized, but with the constraint that no list contained the same word twice and that within a trial no more than two words from the same semantic category occurred in succession.

**Processing load conditions.** Three processing load conditions were created to assess whether variation in processing load would differentially influence storage (i.e., immediate word recall). For each processing load condition, subjects were presented five list lengths (word lists), for a total of 15 recall trials (5 word lists × 3 trials). Each processing load condition included five list lengths that
contained unique randomized orders of words. That is, no condition contained
the same set of words at any given list length or for any given trial within a
given list. Finally, no condition contained a serial progression of word lists that
began with a 3-word list and progressed incrementally to a 7-word list. That is,
each condition included a randomly ordered set of list lengths, none of which
began with a 3-word list and ended with a 7-word list. This procedure was
intended to minimize poorer performance on longer lists by disallowing them
from systematically occurring after shorter lists. All subjects received all list
lengths in each processing load condition. Processing load conditions were pre-
sented in a counterbalanced fashion across subjects.

No-load condition (simple span). Subjects were presented with five word lists
and were asked to recall as many words as possible from each list, regardless
of the order of presentation (i.e., free recall). This simple word span condition
served as an index of storage, independent of any complex processing loads,
against which the next two processing load conditions were compared. This
condition asked for free recall and not sequential recall. We were interested
in estimating simple span apart from any additional processing requirements
associated with retaining the serial position of words (i.e., tagging temporal
location of each word in a list) (see Gillam et al., 1995, for a discussion of the
processes underlying serial recall). Asking for free recall rather than serial recall
also potentially circumvented the potential use of the phonological loop (i.e.,
need for retaining serial order of phonological input). Given the findings of
Gillam et al. (1995) showing that children with SLI and CA controls demon-
strate similar digit spans for free recall (but not serial recall), the children with
SLI and their CA and RS peers might be expected to yield comparable simple
word spans in the present study.

Single-load condition (size processing). Subjects were presented with five word
lists and were asked to recall as many of the words as possible but to reorder
the words upon recall according to physical size of the word referent, “starting
with the smallest thing and ending with the biggest thing.” Children were told
to think of a “typical” example of each of the word referents as they ordered
the words by size. Word lists were constructed with the constraint that no two
immediately adjacent words within a semantic category (see Appendix 1) oc-
curred in a list. The intent here was to create lists in which the word referents
could be ordered by size with relatively minimal “perceptual” confusion among
the items’ relative sizes. This condition served as a single-processing load task
because subjects had to store the words at each list length as well as to process
each word-referent’s physical size relative to each other prior to (or during)
recall. Thus, this condition allowed us to examine the influence of performing
one mental operation on children’s storage capacity. Compared to the simple
span condition, we might expect that this condition should be more difficult for
the SLI children.

Dual-load condition (semantic categorization + size processing). For each of the
five word lists, subjects were asked (a) “to put the words that go together in
some way in little groups” and (b) to order each of the words inside each group according to size, “starting with the smallest thing and ending with the biggest thing.” For each list length and trial, the number of semantic categories into which words could be grouped was limited to two. Word lists again were constructed with the constraint that no two immediately adjacent words within a semantic category could occur in a list. This condition was intended to serve as the most cognitively demanding task because successful recall performance presumably required subjects to perform two mental operations (semantic categorization and size processing), while simultaneously retaining the words at each list length. This condition thus permitted us to measure the influence of performing two mental operations on the storage abilities of the children. Relative to the other two conditions, we predicted that, compared to the normally developing children and themselves, this condition should be more difficult for the SLI children.

To ensure that all word lists would yield a single preferred order of item recall, each word list was piloted in three phases with 20 different graduate students in each phase. A separate pilot study was conducted with a group of normally developing children. These pilot testing procedures were successful in deriving a preferred order of recall for each word list in each condition (see Appendix 2 for description and results of pilot testing).

**Task 2: Off-line sentence comprehension**

**Sentence stimuli.** The sentence comprehension task was identical to the one used by Montgomery (1995b, in press). Two sets of 20 sentences each were created corresponding to a set of linguistically redundant sentences and a set of linguistically nonredundant sentences. Sentences included a range of constructions and vocabulary appropriate to 6- and 7-year-olds’ comprehension and production abilities (e.g., Miller, 1981). Redundant sentences were of four types: (a) sentences containing double marking of number (e.g., “Point to the picture of the three cats”); (b) semantically reversible sentences with a single embedded subject relative clause (e.g., “The girl who is smiling is pushing the boy”); (c) semantically reversible sentences with a double embedded subject and object relative clause (e.g., “The little boy who is standing is hitting the little girl who is sitting”); and (d) active sentences with adjectival/adverbial material modifying the subject and/or object noun (e.g., “The dirty little boy climbs the big fat tree”). The nonredundant and redundant sentences were nearly structurally identical and encoded essentially the same semantic information. The only difference between the sentence conditions was that the redundant cues and modifying adjectival/adverbial lexical items were absent in the nonredundant sentences, thereby making these sentences shorter (e.g., “Point to the picture of the cats,” “The girl smiling is pushing the boy,” “The little boy standing is hitting the little girl sitting,” “The little boy climbs the fat tree”). The mean number of words contained in the redundant and nonredundant sentences was 11.20 and 7.95, respectively.

A high-quality cassette recording of the stimulus sentences was made of the
same male speaker reading each sentence at a normal conversational rate with normal prosodic variation. Across the 40 experimental trials, redundant and non-redundant sentences appeared randomly.

**Picture stimuli.** For each of the 40 stimulus–sentence pairs (i.e., nonredundant and redundant item), four color pictures were created, one matching the stimulus sentence and three foils. Foil pictures differed from the target picture along just one or two relevant semantic dimensions (e.g., size of the sentence’s subject/object, reversed agency of subject and object, color, or number of objects). A stimulus booklet containing 50 pre-experimental pictures (corresponding to the nouns, verbs, adjectives, adverbs contained in the experimental sentences), 6 practice pictures, and 40 experimental pictures was created. Target pictures appeared equally often in each quadrant on the stimulus page.

**Task 3: Real-time sentence processing**

We employed a word monitoring task (i.e., word recognition reaction time [RT] task) to assess children’s real-time sentence processing. A word monitoring task is regarded as a highly sensitive measure of listeners’ ability to construct immediately (automatically and unconsciously) a semantic–syntactic representation of sentential input as they process the input word by word (Tyler, 1992). The results of the word monitoring task would enable us to examine the time course underlying children’s construction of sentence meaning and to test the hypothesis that an accumulation of sentential input might have a deleterious effect on the sentence-medial or sentence-final lexical processing of SLI children.

**Target words and sentences.** Sentence-embedded target words included 42 easily depictable, highly familiar, monosyllabic nouns (e.g., Moe et al., 1982). Half of the words began with a stop consonant, and the other half began with a nonstop.

There were 84 sentence pairs containing vocabulary and sentence structures appropriate to 6- and 7-year-olds’ comprehension and production abilities (e.g., Miller, 1981). The first sentence in each pair functioned as a topic sentence for the second (test) sentence. A topic sentence was included to provide subjects with a brief context to facilitate processing of the test sentence. The first sentence ranged from 4 to 9 words ($M = 5.9$), and the second ranged from 8 to 14 words ($M = 11.8$). An equal number of target words occurred in the 5th, 7th, or 10th word position of the test sentence. Varying target word location permitted tracking the time course of comprehension. All words preceding the target were acoustically dissimilar to the target, thereby preventing acoustically based false alarms. A total of 12 catch trials were constructed in which a target word did not appear in the test sentence to identify subjects with an impulsive response style.

**Stimulus generation procedures.** The stimulus sentences and a list of individual target words were read aloud by an adult male native speaker of American English. High-quality recordings of all materials were made in a sound-treated booth. Sentences were read at a normal rate with normal prosodic variation;
target words were read in word list fashion. Prior to waveform editing, each recorded sentence was low-pass filtered (4.5 kHz), digitized (10 kHz), and stored on disk. Each digitized audio waveform was edited interactively using the ASYST software package to identify the acoustic onsets and offsets of the sentence-embedded target words. The edited stimuli were stored on disk and later played out (10 kHz) and low-pass filtered (4.5 kHz) from a PC laboratory computer to the subject. A 500 Hz timing pulse was synchronized to begin at the onset of the first sentence. The timing tone, inaudible to the subject, was used during the experiment to trigger an external clock that was used to measure the subject’s word recognition RT. For each sentence, its corresponding isolated target word was low-pass filtered (4.5 kHz), digitized (10 kHz), and stored on disk. Each target word was later played out (10 kHz) and low-pass filtered (4.5 kHz) to the subject 1 second prior to its corresponding sentence.

Auditory detection RT task

Because some SLI children show slower RT than age controls even on tasks having little to do with language (e.g., Edwards & Lahey, 1996; Hughes & Sussman, 1983; Lahey & Edwards, 1996), it seemed important to include a simple auditory detection RT task. This task served as an index of motor response and auditory reception/sensation time independent of linguistic processing. If the SLI children were slower than the age controls on this simple RT task, each group’s mean RT could be used as a covariate during data analysis of the word recognition RT task. Subjects were instructed to press a response pad as quickly as possible in response to the onset of an imperative stimulus (1 s, 2 kHz pure tone) which followed a warning tone (500 ms, 500 Hz tone). Again, both response accuracy and speed were stressed to the subject. Following 5 live-voice practice trials (i.e., examiner producing a “low beep” followed by a “high beep” and demonstrating the button press on the high beep) and 15 computer-delivered practice trials, 36 experimental RT trials (interstimulus interval) between warning and test tones varying randomly between 1.5 and 3 seconds were presented. Subjects received constant encouragement and praise as well as necessary reminders “to stay alert” during the task. A mean auditory detection RT was calculated for each subject. This task provided subjects with practice and familiarity with the word recognition RT task and was intended to enhance intrasubject RT stability. This task always preceded the word recognition RT task.

Two counterbalanced orders of the word recognition RT task were created. Two counterbalanced orders of the experimental tasks also were created. Equal numbers of subjects received one of the two orders.

PROCEDURES

Working memory task

Subjects were tested individually in a sound-treated acoustic booth. Wearing headphones, subjects received the word lists and the sentences at a comfortable
listening level. They were told in general that (a) they were going to hear a man saying some lists of words and that they needed to listen carefully because (b) they needed to repeat as many of the words from each list as they could remember. They were told that for some of the lists of words they would be asked “to rearrange the words in their mind in a special way” before or as they repeated them.

**Pretesting.** Prior to the experiment proper, subjects were administered three brief pretests to assess their ability to understand the processing demands inherent in the single-load and dual-load conditions. In each pretest, subjects were presented with highly familiar, monosyllabic words (not the experimental words) in lists varying from 3 to 4 words. Three demonstration trials and three practice trials preceded eight pretest trials. During the demonstration trials, the examiner used his hands and fingers to illustrate the idea of grouping and/or ordering the words by semantic category and/or by physical size. In the first pretest, a semantic categorization task, subjects were asked to group those words that went together in some way next to each other as they repeated the words back to the examiner. In the second pretest, a size processing task, subjects were asked to reorder the words “starting with smallest thing and ending with the biggest thing” as they recalled as many words as they could. In the third pretest, semantic categorization + size processing, subjects were told that they “would be doing both things with each list of words.” They were instructed (a) to put the words that went together in some fashion in little groups and (b) to order the words in each little group from the smallest thing to the biggest thing. Subjects were allowed one repetition of the input and two opportunities to recall the word list. To advance to the experimental proper, subjects had to respond correctly on 6 of 8 trials (75% accuracy) on the first two pretests (see Data Analysis for definition of a correct response). These procedures thus facilitated our interpretation that any observed group differences on the experimental task would be attributable to differences in working memory and not to a lack of understanding the task.

**Experiment proper.** Subjects were again instructed to listen carefully to a man saying some lists of words. They were encouraged to repeat as many of the words as they could in the order that was required for each condition. Throughout testing, subjects received constant encouragement and praise and, as needed, reminders to “listen carefully” and “stay focused.” Children were provided one repetition of any word list if needed and two opportunities for recall. Children’s second production was scored.

**Off-line sentence comprehension task**

Subjects received the 40 experimental sentences via headphones at a comfortable listening level. While listening to each sentence, subjects were shown an array of four pictures. After hearing each sentence, they were asked to point to the picture corresponding to the sentence. Subjects were allowed one additional presentation of each stimulus sentence. Subject responses were scored as correct
or incorrect. Prior to experimental testing, a pretest was administered to assess subjects’ knowledge of the nouns, verbs, adjectives, and adverbs contained in the experimental sentences. All subjects performed with 100% accuracy.

Real-time sentence processing task
Sentence stimuli were delivered to the subject binaurally via headphones at a comfortable listening level. A total of 4 live-voice practice items and 12 computer practice items preceded the experimental items. Subjects were told to listen to some “short stories” and to push a response pad as quickly as they could as soon as they heard the target word in the “story.” Both response accuracy and speed were stressed to subjects in the instructions and practice trials as well as during the experimental task as needed. Prior to stimulus presentation, subjects were shown a picture of the target word by a second experimenter who sat with the subject in the test booth. Throughout testing, subjects received constant encouragement and praise and, as needed, reminders to “stay focused.” These forms of nonspecific feedback were intended to maximize both motivation and RT performance. Subject responses stopped the clock. Responses occurring prior to target word onset (i.e., negative RT) were scored as false alarms, and failures to respond were scored as misses. Subjects’ RT were computed by a custom-written program and stored on the computer.

Data analysis
For the working memory task, the dependent variable in each processing load condition was word span, defined as the longest list length for which a subject correctly responded on at least two of the three trials. If the subject produced three different spans at a given list length (e.g., stimulus: 5-word list length; responses: 3 words recalled, 4 words recalled, 5 words recalled), word span was defined as the middle span (e.g., 4 words). A correct response in the no-load condition was defined as the subject recalling as many words as possible at each list length, regardless of presentation order. In the single-load condition, a correct response was defined as the greatest number of words properly (sequentially) ordered by size during recall. In the dual-load condition, a correct response was defined as the greatest number of words properly grouped by semantic category and properly (sequentially) ordered by size during recall. Thus, subjects produced three different word spans: simple span (no-load condition), size span (single-load condition), and semantic category + size span (dual-load condition). Word spans were used in each of the analyses. The dependent variable in the off-line comprehension task was the number of sentences correctly comprehended of each sentence type. Word recognition RT was the dependent variable in the on-line processing task.

Predictions
We predicted that the SLI children should demonstrate poorer recall in both the single-load and dual-load conditions relative to the age controls and possibly the younger children. For the off-line comprehension task, we hypothesized that
the SLI children relative to the control children should comprehend fewer redundant sentences than nonredundant sentences. We also expected to see a significant correlation between performance on the comprehension task and the dual-load condition. Even though real-time processing tasks minimize the use of working memory relative to off-line tasks (Tyler, 1992), we hypothesized that, if working memory limitations in SLI children do affect their immediate sentence processing, one of two primary word recognition RT patterns should emerge. The SLI children should demonstrate either no word recognition RT advantage for words occurring in the middle and/or at the end of the test sentences or progressively (significantly) slower RT at each successive word position. This prediction derives from the assumption that, if SLI children have difficulty retaining information occurring in the topic sentence and/or at the beginning of the test sentence, they should demonstrate no RT advantage for words appearing later in the test sentence. Given that listeners develop an immediate interpretation of each utterance they hear (Tyler, 1992), we assumed that the children should store a representation of the topic sentence, which, in turn, should facilitate the processing of the test sentence. In addition, a significant correlation between working memory and word recognition RT might be expected.

RESULTS

Working memory task

The analyses focused on determining whether variation in processing load would have a differential influence on storage, both between and within groups. As indicated by Figure 1 and Table 2, the subject groups appeared to perform
Table 2. Mean longest word span produced by the subjects with SLI and the CA and RS controls under each processing load condition: no load (simple word span), single load (size processing), and dual load (semantic + size processing)

<table>
<thead>
<tr>
<th>Processing load condition</th>
<th>No load</th>
<th>Single load</th>
<th>Dual load</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>4.91</td>
<td>4.66</td>
<td>3.83</td>
</tr>
<tr>
<td>SD</td>
<td>.51</td>
<td>.48</td>
<td>.39</td>
</tr>
<tr>
<td>Range</td>
<td>4–6</td>
<td>4–5</td>
<td>3–4</td>
</tr>
<tr>
<td>CA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>5.08</td>
<td>4.83</td>
<td>4.83</td>
</tr>
<tr>
<td>SD</td>
<td>.37</td>
<td>.83</td>
<td>.57</td>
</tr>
<tr>
<td>Range</td>
<td>4–6</td>
<td>4–6</td>
<td>4–6</td>
</tr>
<tr>
<td>RS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>4.66</td>
<td>4.58</td>
<td>4.00</td>
</tr>
<tr>
<td>SD</td>
<td>.49</td>
<td>.52</td>
<td>.60</td>
</tr>
<tr>
<td>Range</td>
<td>4–5</td>
<td>4–5</td>
<td>3–4</td>
</tr>
</tbody>
</table>

comparably in the no-load and single-load conditions. However, in the dual-load condition, the CA children seemed to outperform both the SLI and RS children. The SLI children and RS children appeared to perform similarly to one another in each condition. These impressions were supported by the results of a Group (3) × Processing Load (3) repeated measures ANCOVA. Nonverbal IQ was used as the covariate to adjust for any influence that the lower IQ of the SLI children may have played between the groups’ performances. Results revealed significant effects for group, $F(2, 98) = 8.88, p < .001$, and processing load, $F(2, 98) = 14.26, p < .001$, and a significant Group × Processing Load interaction, $F(4, 98) = 2.47, p < .05$.

Between-group post hoc ANCOVAs (using IQ as the covariate) on each of the task load conditions showed that the subject groups indeed performed similarly under the no-load condition, $F(2, 32) = 2.65, p > .05$, and the single-load condition, $F(2, 32) = .445, p > .05$. Under the dual-load condition, a group difference emerged, $F(2, 32) = 9.97, p < .001$. The CA children outperformed both the SLI and RS children (Tukey HSD analysis, $p < .05$). The SLI and RS groups performed comparably (Tukey HSD analysis, $p > .05$).

As suggested by Figure 1, the recall performance of the CA children was unaffected by processing load, whereas the recall performance of the SLI and RS children declined under the dual-load condition. A series of within-group post hoc ANOVAs confirmed these impressions. First, no significant condition effect obtained for the CA children, $F(1, 33) = .673, p > .05$. By contrast, a significant condition effect was found for the SLI children, $F(1, 33) = 17.57, p < .001$, and RS children, $F(1, 33) = 5.45, p < .001$. Both groups demonstrated poorer recall under the dual-load condition relative to each of the other two
conditions (Tukey HSD analysis, \( p < .05 \)). No other condition comparisons reached significance.

To help explain the nature of the SLI and RS children’s poorer performance on the dual-load condition, an error analysis was performed on the number of errors related to semantic categorization versus size processing for the 5-, 6-, and 7-word lists (i.e., for above-span word lists). Inspection of the errors (pooling across word lists) revealed that the overwhelming majority of recall errors by the children with SLI (80%) and the RS children (83%) related to difficulty in arranging the words by size and/or in remembering a word within a given semantic category rather than errors of semantic categorization (i.e., ordering words by size without regard to category). Specifically, while the children were able to establish two semantic categories (albeit one category may have contained only one item), their greatest difficulty was in ordering the words by size and/or in remembering all of the words within each category.

**Off-line comprehension task**

The SLI children relative to the CA and RS children yielded a different performance pattern across sentence type conditions. As can be seen from Figure 2 and Table 3, sentence type had no differential influence on the CA and RS children’s comprehension, whereas the children with SLI appeared to perform more poorly on the redundant sentences compared to the nonredundant sentences. Results of a Group (3) × Sentence Type (2) repeated measures ANOVA supported these observations. Results revealed significant effects for group, \( F(2, 66) = 23.85, p < .001 \), and sentence type, \( F(1, 66) = 37.36, p < .001 \), and a significant Group × Sentence Type interaction, \( F(2, 66) = 19.62, p < .001 \).

Separate post hoc between-group ANOVAs were computed to examine which groups differed from one another across conditions. The SLI–CA group comparison revealed effects for group, \( F(1, 44) = 46.29, p < .001 \), and sentence type, \( F(1, 44) = 31.13, p < .001 \), as well as a Group × Sentence Type interaction, \( F(1, 43) = 30.99, p < .001 \). The SLI children comprehended fewer nonredundant sentences, \( t(22) = −2.89, p < .001 \), and redundant sentences, \( t(22) = −7.03, p < .001 \). The SLI–RS group comparison also revealed effects for group, \( F(1, 44) = 18.44, p < .001 \), and sentence type, \( F(1, 44) = 38.87, p < .001 \), as well as a Group × Sentence Type interaction, \( F(1, 44) = 24.52, p < .001 \). However, whereas the SLI children and their RS peers comprehended a comparable number of nonredundant sentences, \( t(22) = −1.29, p > .05 \), the SLI children comprehended fewer redundant sentences, \( t(22) = −6.12, p < .001 \). Finally, the CA–RS group comparison revealed a significant group effect, \( F(1, 44) = 4.87, p < .05 \); the CA children demonstrated better performance than the RS children overall. No sentence type effect or Group × Sentence Type interaction obtained.

Within-group analyses for sentence type showed that the SLI children comprehended significantly fewer redundant sentences than nonredundant sentences, \( t(11) = 6.92, p < .001 \). By contrast, the CA children performed comparably in the nonredundant and redundant conditions, \( t(11) = 1.61, p > .05 \), as did the RS children, \( t(11) = 1.37, p > .05 \).

To help clarify the nature of the poorer performance of the SLI children, a
Figure 2. Mean number of nonredundant and redundant sentences comprehended by children with SLI and the CA and RS children. Error bars represent the standard error of the mean.

Table 3. Mean number of sentences comprehended by the children with SLI and the CA and RS controls under each sentence type condition: Nonredundant and redundant

<table>
<thead>
<tr>
<th>Sentence type</th>
<th>Nonredundant</th>
<th>Redundant</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>17.0</td>
<td>13.9</td>
</tr>
<tr>
<td>SD</td>
<td>.99</td>
<td>1.62</td>
</tr>
<tr>
<td>Range</td>
<td>16–19</td>
<td>11–17</td>
</tr>
<tr>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>18.1</td>
<td>17.7</td>
</tr>
<tr>
<td>SD</td>
<td>.67</td>
<td>.89</td>
</tr>
<tr>
<td>Range</td>
<td>17–20</td>
<td>16–20</td>
</tr>
<tr>
<td>RS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>17.5</td>
<td>17.2</td>
</tr>
<tr>
<td>SD</td>
<td>.90</td>
<td>1.02</td>
</tr>
<tr>
<td>Range</td>
<td>16–19</td>
<td>17–19</td>
</tr>
</tbody>
</table>

pattern analysis on the number of correct responses was performed on each of the four different sentence constructions for the redundant sentences. The SLI children obtained the following mean number correct for each of the following sentence constructions: double marking of number, 4.8; single embedded relative clause sentences, 3.2; double embedded relative clauses, 2.5; and sentences with extra verbiage, 3.4. The following mean number correct for each of the
following sentence constructions was obtained by the CA children: double marking of number, 4.8; single embedded relative clause, 4.3; double embedded relative clauses, 4.4; and sentences with extra verbiage, 4.2. The RS children obtained the following mean number correct for each of the following sentence constructions: double marking of number, 4.8; single embedded relative clause, 4.2; double embedded relative clauses, 3.9; and sentences with extra verbiage, 4.3.

**Auditory detection RT task**

The groups did not differ in their accuracy of responding (i.e., number of valid RT responses), $F(2, 33) = 1.78, p > .05$. The SLI and CA children yielded a hit rate of 96% and 97%, respectively, while the RS children produced a hit rate of 94%. A mean RT was calculated for each subject. All outlier responses, defined as RTs falling ±2 SDs from a subject’s mean RT, were eliminated (Fazio, 1990). The children produced relatively few outlier responses (SLI: 3.1%, CA: 2.9%, VM: 3.3%). Missing RT values were replaced with the child’s mean RT. Overall, the groups yielded statistically comparable auditory detection RT, $F(2, 33) = 1.85, p > .05$ (SLI: 434 ms, CA: 353 ms, RS: 389 ms). Thus, this variable was not used as a covariate in the sentence processing task analysis.

**Real-time sentence processing task**

Again, the groups produced similar hit rates (i.e., number of valid RT trials), $F(2, 33) = 2.08, p > .05$. The SLI children attained a hit rate of 91%, CA children attained a hit rate of 90%, and RS children achieved an accuracy rate of 88%. Subject groups also produced very few responses to catch trials and did not differ on this variable, $F(2, 33) = .389, p > .05$.

For each subject, a mean RT was calculated for each word position. Individual RTs that fell ±2 SDs from a subject’s mean RT were eliminated as outlier responses. Relatively few responses were excluded for this reason (SLI: 3.3%, CA: 3.4%, RS: 3.7%). In the case of any missing RT values due to misses or false alarms, an appropriate mean RT for those children at a given word position was inserted into that child’s data set. Subjects’ mean RTs were used in all subsequent analyses.

Results of a Group (SLI, CA, RS) × Word Position (5th, 7th, 10th) repeated measures ANOVA revealed main effects for group, $F(1, 99) = 16.48, p < .001$, and word position, $F(1, 99) = 22.17, p < .001$. No significant Group × Word Position interaction obtained, $F(2, 99) = .682, p > .05$. As can be seen from Figure 3 and Table 4, the group effect can be explained by the fact that the CA children had faster word recognition RT overall than either the SLI children, $F(1, 66) = 13.17, p < .001$, or the RS children, $F(1, 66) = 3.88, p < .05$, and by the fact that the RS children yielded faster RT overall than the children with SLI, $F(1, 66) = 14.22, p < .001$. Post hoc between-group ANOVAs on RT at each word position were carried out to evaluate the effect of word position on RT between each group. Relative to the CA children, the SLI children yielded slower RT in both the 5th position, $F(1, 22) = 13.67, p < .001$, and 7th position, $F(1, 22) = 6.57, p <$
Table 4. Mean word recognition RT by word position for the children with SLI and the CA and RS children

<table>
<thead>
<tr>
<th>Word position</th>
<th>SLI</th>
<th>CA</th>
<th>RS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3rd</td>
<td>7th</td>
<td>10th</td>
</tr>
<tr>
<td>Grand mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>500</td>
<td>365</td>
<td>339</td>
</tr>
<tr>
<td>SD</td>
<td>114</td>
<td>115</td>
<td>112</td>
</tr>
</tbody>
</table>

.01, but not in the 10th position. Relative to the RS children, the SLI children produced significantly slower RT in the 5th position only, $F(1, 22) = 10.31, p < .001$. For the CA–RS comparisons, no reliable RT differences were found at any word position. Table 4 displays subject groups’ mean RT by word position.

Figure 3 also reveals that word recognition RT generally became faster across word position for all the children. To assess the effect of word position on each group’s RT, separate one-way ANOVAs and post hoc Tukey HSD analyses ($p < .05$) were performed. The SLI children demonstrated faster word recognition
RT across word position, $F(1, 33) = 6.27, p < .01$. Their RT was significantly faster by the 7th word position relative to the 5th word position. Their RT to words appearing in the 10th position, however, was no faster than words appearing in the 7th position. Likewise, the CA children demonstrated progressively faster RT across word position, $F(1, 33) = 9.22, p < .001$. They showed faster RT in the 7th position relative to the 5th position, but no RT advantage in the 10th position relative to the 7th position. The RS children also demonstrated a significant word position effect, $F(1, 33) = 13.31, p < .001$. They yielded significantly faster RT in the 10th position relative to 7th position, but no RT advantage by the 7th position relative to the 5th position.

**Relation between working memory and off-line and real-time comprehension**

The correlation between performance on the different conditions of the working memory task and performance on the sentence comprehension task was assessed using a series of Pearson product moment correlations ($p < .01$) for the SLI children and the control children (CA and RS groups combined) separately. For the SLI children, a significant negative correlation ($−.43$) was found between performance on the dual-load condition and total score on the sentence comprehension task. No other significant correlations emerged. For the control children, a significant positive correlation was found between performance on the dual-load condition and total score on the comprehension task ($+.47$); no other correlations reached significance.

To assess the relation between working memory and on-line language processing, a series of Pearson product moment correlations ($p < .01$) was computed between performance on the dual-load processing task and performance on the word recognition RT task using overall word recognition RT and RT at each word position as the dependent variables. No significant correlations were found for either the SLI children or the control children (CA and RS groups combined).

**DISCUSSION**

**Working memory task**

The prediction that SLI children should demonstrate poorer storage under both processing load conditions relative to their normally developing counterparts was partially supported. Counter to our prediction, the SLI children and their CA peers performed similarly in both the no-load and single-load conditions. In the dual-load condition, however, the SLI children performed worse, a finding that was consistent with our expectations. The SLI and RS children, on the other hand, performed comparably to each other in each of the three conditions. Whereas the CA children’s recall was unaffected by processing load, the SLI and RS children’s recall in the dual-load condition decreased relative to the other two conditions. By contrast, recall was comparable for both the SLI and RS groups in the no-load and single-load conditions. These findings are consistent with those reported by Montgomery (in press). Significantly, the no-load
condition findings suggest that, in the absence of any additional processing requirements, the SLI children have comparable “simple” storage capacity to their CA and RS peers. Because this condition asked for free recall and not serial recall and because it presumably minimized the potential role of the phonological loop (i.e., obviated demand to retain the input in serial order), this finding should not be too surprising (Gillam et al., 1995).

Relative to the SLI and the youngest children, the CA children evidently experienced no trade-off between storage and processing, even in the most demanding processing condition. The CA children apparently were more efficient in managing both the storage and processing functions of working memory than their SLI and younger peers. This finding runs counter to the expectations of the working memory model of Daneman and colleagues, which assumes that storage should suffer in the face of increasing processing demands. The strong performance of the CA children, however, may be a reflection of this specific task. The task simply may not have sufficiently taxed these children’s working memory resources. Apparently, these children were able to perform both concurrent mental operations very efficiently, thereby freeing up more of their working memory resources for storage.

The fact that the SLI and RS children’s recall was no worse in the single-load condition compared to the no-load condition suggests that the SLI children had some ability to coordinate storage and processing. Under the single-load condition, the working memory resources of the SLI children evidently were not exceeded by having to store the input words and order them by size at the same time. The SLI children, however, did demonstrate decreased storage under the most demanding condition relative to themselves and the CA children – a finding that was consistent with our predictions. These results suggest that, relative to their CA peers, the SLI children had less functional working memory capacity yet comparable capacity to their RS peers. For these SLI children, there indeed appeared to be a cost to storage in the context of having to perform a complex processing task. On Daneman and colleagues’ view of working memory, the SLI children apparently allocated their working memory resources more to processing and less to storage. These children’s reduced functional working memory capacity is, significantly, in direct contrast to their larger “simple” storage capacity, as indexed by their similar recall to the CA children in the no-load condition.

How might the poorer recall of the SLI children under the dual-load processing condition relative to the CA children be explained? The SLI children may have possessed poorer specific semantic knowledge and basic semantic processing abilities. However, this possibility is not likely. First, the words were highly familiar and easily categorizable (as indicated by pilot testing). Second, SLI children of this age have been shown to use semantic categorization (e.g., Kail & Leonard, 1986). Third, the results of pretesting indicated that all of the children were able to categorize according to semantic feature, thus suggesting that these SLI children had at least basic semantic categorization abilities. Nonetheless, despite these children’s possessing the necessary specific semantic knowledge and basic processing abilities, they may not have been as efficient in deploying their knowledge and processing abilities in the face of having to
perform two concurrent mental operations. The dual demands of categorizing words and ordering them by size may very well have disrupted their efforts at efficient semantic processing, thereby leading to a functional decrease in storage. Support for this interpretation comes from the finding that the majority of the errors by the SLI children (and RS children) were related to their forgetting some of the words from one or both categories and/or improperly ordering them by size within an established category. At the same time, efficient semantic processing during a complex processing task presumably would be facilitated in those children who possess greater lexical knowledge and a more extensive network of lexical knowledge. Relative to the CA children, it is possible, then, that the reduced lexical knowledge and less elaborate network of lexical knowledge of the SLI children (Kail & Leonard, 1986) may have contributed to their poorer performance. That is, even though the experimental words were highly familiar, they may not have remained in a sufficiently highly activated state long enough for these children to complete both mental operations (i.e., words could not be adequately rehearsed and refreshed in short term memory), resulting in decreased functional storage capacity.

Differences between the SLI and CA children in various short term memory processes also might have contributed to some degree to the inferior recall of the SLI children. For instance, SLI children have been shown to scan short term memory more slowly than their age peers (Sinninger et al., 1989). Thus, it is possible that the SLI children may have been slower scanning short term memory for all of the to-be-remembered words while also attempting to complete both mental operations. They also might have been unable to complete both mental operations before the words began to fade from short term memory (e.g., Gathercole & Baddeley, 1990a). Difficulty maintaining words in short term memory, as opposed to difficulty forming accurate phonological/lexical representations or possessing less distinct phonological/lexical representations (e.g., Edwards & Lahey, 1998), might also be a likely possibility. If forming accurate phonological representations of the words were the problem, the SLI children should have demonstrated poorer recall across conditions. This pattern was not observed, however. Similarly, access to less phonologically distinct lexical representations is not very likely, given that the stimuli were highly familiar, monosyllabic words (not nonsense words) and that these children performed well in both the no-load and single-load conditions. Problems with verbal rehearsal and motor planning/execution also are not strong possible sources of difficulty for these SLI children. If they were, the SLI children should have performed equally poorly in all of the processing conditions. Again, this pattern was not found. In support of this claim, several investigators indeed have shown that, relative to age peers, SLI children demonstrate comparable verbal rehearsal processes (Gathercole & Baddeley, 1990a; Kail & Leonard, 1986; Sinninger et al., 1989) and articulatory planning, speed, and execution (Edwards & Lahey, 1998; Gathercole & Baddeley, 1990a; Montgomery, 1995a; Stark & Montgomery, 1994). Taken on the whole, the poorer recall of the SLI children in the dual-load condition appears to be related primarily to their difficulty maintaining the words in short term memory, presumably because fewer working memory resources were allocated to storage and a greater amount were allocated to processing. Given
the highly semantic nature of this condition, a less elaborate network of lexical knowledge leading to less efficient semantic processing abilities (resulting in inefficient rehearsal and item decay) by the SLI children may also have been a contributing factor.

Off-line sentence comprehension and its relation to working memory

The SLI children performed worse overall than their CA and RS peers. Relative to the CA children, the SLI children comprehended fewer redundant and nonredundant sentences. Relative to their RS peers, the poorer performance of the SLI children was attributable to their comprehending fewer redundant sentences. Moreover, the SLI children comprehended fewer redundant versus nonredundant sentences relative to themselves. By contrast, neither the CA nor the RS children showed superiority for one sentence type over another. The poorer comprehension of the redundant sentences by the SLI children relative to the younger children replicates the findings reported by Montgomery (1995b, in press). The present findings also extend the Montgomery findings by demonstrating that, relative to CA children, SLI children displayed sentence comprehension impairment, regardless of sentence length. These latter findings suggest that the poorer comprehension of the SLI children was likely due to both incomplete syntax knowledge and reduced functional working memory capacity.

The poorer comprehension of redundant sentences by the SLI children relative to their RS peers was not attributable to poor sentence-level syntax knowledge but instead to difficulty managing the increased demands on verbal working memory. Support for this interpretation comes from the fact the SLI and RS children performed similarly on the nonredundant sentences, which were essentially identical in syntactic form to the redundant sentences. As sentence length increased, the demands for storing just-processed input while simultaneously processing new, incoming information also increased. The SLI children’s poorer comprehension of redundant sentences thus appears to reflect, at least in part, their difficulty with having to store greater amounts of earlier input while at the same time rapidly computing the syntactic—semantic representation of new, incoming information. This interpretation is supported by the finding that the children with SLI had the greatest difficulty comprehending sentences containing double relative embedded clauses (i.e., “The girl who is crying is pointing to the boy who is laughing”), followed by single relative embedded clause items (i.e., “The boy who is sitting behind the girl is very big”) and sentences with extra verbiage (“The little old blue car is going to hit the great big fast speeding train”). The CA and RS children, on the other hand, showed roughly comparable performances across the different sentence constructions. In Dane-man and associates’ view of the relation between working memory and comprehension, the difficulty of the SLI children processing the relative clause sentences may have arisen because, by the time they had to process the sentence-medial or sentence-final phrase/clause, the thematic representation associated with the subject noun phrase encountered at the beginning of the sentence had been forgotten. Thus, these children may have had trouble establishing the syntactic-thematic link between the initial noun phrase and the sentence-medial
or sentence-final phrase/clause. It seems reasonable to assume that a greater amount of these children’s working memory resources may have been allocated to the syntactic–semantic processing of the later-occurring input, leaving fewer resources for storing the sentence-initial information. Or, the children may have allocated most of their resources to storing the initial part of the sentence, leaving fewer resources for processing later-occurring information. The positive correlation between subjects’ performance on the dual-load processing condition and performance on the sentence comprehension task lends further support to the interpretation of some kind of storage/processing trade-off. Assuming sentence comprehension entails coordination between storage and the simultaneous operation of multiple comprehension processes, it should come as no surprise that a correlation was found. Although the correlations were only $-0.43$ (SLI group) and $+0.47$ (control groups combined), these are in the range of correlations reported by Daneman and Merikle (1996) for a wide variety of working memory and comprehension tasks used with adults. It is noteworthy that it was the combined functions of storage and processing represented by the more complex processing condition – not storage alone (simple span) – that correlated with sentence comprehension in the present study. This finding is consistent with reports by Daneman and colleagues demonstrating that the combined functions of storage and processing are better predictors of comprehension than simple span. Finally, while a correlation between the dual-load condition and sentence comprehension was found, it is important to point out that other sources of variance (e.g., rapid phoneme identification, trace decay) also may have contributed in some way to the comprehension difficulties of these SLI children.

The fact that the SLI children did not differ from the CA and RS children with respect to storage capacity, along with the finding that storage alone was not associated with sentence comprehension, may at first glance seem inconsistent with the interpretation offered by Montgomery (1995b), which proposed that a capacity limitation in working memory is associated with these children’s sentence comprehension impairment. This apparent discrepancy lies in the nature of the estimates of storage capacity and differences in theoretical views of capacity that were used in each study. In the Montgomery study, capacity, defined from within the framework of Baddeley’s phonological loop model, was estimated by the ability to repeat nonsense words varying in length. Repetition of nonsense words is generally regarded as a “purer” measure of phonological memory capacity to the extent that successful repetition is not aided by lexical knowledge, as is the case when recalling real words. Asking subjects to recall as many real words as possible (as required in this study) presumably does not stress phonological memory capacity in the same way that the repetition of nonsense words does (i.e., demand for retaining serial order of phonological input). In the present study, “simple” capacity was indexed by the number of real words reproduced during free recall (not serial recall), and functional capacity was indexed as the number of words recalled while simultaneously engaged in processing (i.e., index of the ability to allocate resources to both storage and processing). It is likely, however, that SLI children’s apparent limitation in phonological memory capacity (Gathercole & Baddeley, 1990a; Montgomery, 1995b), along with their reduced functional working memory capacity, jointly
Montgomery: Working memory and sentence processing

contributed in varying degrees to some of their sentence comprehension problems.

The fact that the SLI children comprehended fewer redundant sentences than their RS peers yet performed comparably to them on the working memory task must be reconciled in some fashion. The reconciliation seems to lie in the nature of the task and how intrinsic information processing abilities of the child interact with task (Edwards & Lahey, 1998; Lahey & Bloom, 1994). In the working memory task, the SLI and RS children were equally capable of coordinating to some extent the joint demands of storage and processing. The comprehension task, however, evidently exceeded the working memory resources of the SLI children. Children were required to store and comprehend each input sentence and choose the one picture (from an array of four) that matched the input sentence. At the same time, they also presumably needed to generate a linguistic representation of each foil picture, store each of these representations, and then compare each representation with the representation of the input sentence and with each picture. These additional processing requirements likely overtaxed the working memory resources (and general processing capacity) of the SLI children and may have been responsible for their especially poor comprehension of the redundant sentences (Montgomery, in press).

Real-time sentence processing and its relation to working memory

The RT pattern produced by the SLI children was not consistent with a hypothesis that slower word recognition might be related to limited functional working memory capacity. On this account, it could be predicted that the word recognition RT of the SLI children should either remain flat across all word positions or become progressively/significantly slower at each successive word position. This prediction follows from the assumption that, if these children had difficulty retaining information heard in the topic sentence and/or from the beginning of the test sentence, their RT to words appearing in the middle and end of the test sentence would not be facilitated. Instead, the RT pattern of the SLI children is consistent with the interpretation that they are slower at linguistic retrieval and/or evaluating the linguistic properties of incoming words relative to prior context (Montgomery & Leonard, 1998; Montgomery et al., 1990; Stark & Montgomery, 1995). This interpretation is supported by the fact that there were main effects for both group and word position but no group by word position interaction. Even though the SLI children yielded slower word recognition RT at certain word positions relative to the CA and younger children, all the children produced faster RT at later word positions. These findings suggest that, while the SLI children were slower to process the input relative to their normally developing peers, their sentence processing, like that of the normally developing children, was indeed facilitated by an accumulation of sentential information.

The present findings are generally consistent with those of Stark and Montgomery (1995) and Montgomery et al. (1990). Stark and Montgomery showed that the word recognition RT of children with SLI was slower than that of their CA peers. But they also showed that CA and SLI children’s word recognition RT was faster by the middle of the sentence, although by the end of the sentence
the facilitatory effects of additional input had begun to diminish. Like the find-
ings of Stark and Montgomery, the present findings indicate that SLI children,
like their CA peers, are able to construct a “well-developed” sentence interpreta-
tion by about the middle of a sentence, and that additional information does not
appear to significantly facilitate faster processing. In fact, the present findings
show that by the end of the sentence the SLI and CA children have comparable
word recognition RT. While the present findings suggest that the lexical pro-
cessing and development of a sentence-level representation by the SLI children
benefited from an accumulation of input up to a certain point, they also indicate
that these children were slower to begin to construct such a representation. This
interpretation is supported by the fact that the SLI children’s RT in the 5th word
position was slower than the RT for both the CA and younger children. This
finding is consistent with the findings reported by Montgomery et al. (1990).

It might be argued that the word recognition RT of the SLI children was no
faster in the 10th position relative to the 7th position because they did not retain
sufficient amounts of earlier information to facilitate word recognition at the
end of the sentence. This argument does not hold for a couple of reasons. First,
the CA controls and the SLI children demonstrated the same RT pattern. Sec-
ond, the absence of a correlation between working memory and word recogni-
tion RT (i.e., overall RT and RT for 10th word position) runs counter to a
working memory deficit interpretation. Together, these findings provide evi-
dence for the tentative conclusion that the immediate processing of sentential
input by SLI children is not hindered by their more-limited functional working
memory capacity. In fact, working memory does not appear to be involved, or
at least not very strongly, in the real-time sentence processing of children with
SLI or their normally developing peers. Unless or until evidence to the contrary
emerges, the slower real-time processing of children with SLI appears to be
primarily related to inefficient linguistic retrieval operations.

It is possible that a relation between working memory and on-line sentence
processing was obscured because the sentences used in the off-line task were
more difficult to process than the ones used in the on-line task. Given that
the sentences used in each task were constructed using similar constraints (i.e.,
vocabulary, syntactic–semantic structure), this possibility is not very likely. In
addition, it seems reasonable to assume that, if there were differences in sen-
tence complexity, any small differences would be offset by the presumably in-
creased working memory load associated with having to process sentence pairs
in the on-line task (assuming the topic sentence was indeed processed and
stored) as opposed to single sentences in the off-line task. Thus, it seems more
likely that no relationship between working memory and on-line sentence pro-
cessing was observed because the sentences were well within the linguistic grasp
of all the children, thereby allowing them to effectively deploy a range of se-
matic–syntactic processing operations. The linguistic processing demands in-
herent in these sentences apparently did not reach the limits of the real-time
working memory resources of any of the children, including the SLI children.
These findings and interpretations are consistent with Tyler’s (1992) claim that
on-line tasks assess listeners’ unconscious mental representations and automatic
operations, independent of significant working memory demands.
SUMMARY

The findings from this study provide further evidence for a processing deficit account of the off-line sentence comprehension problems observed in some SLI children (Montgomery, 1995b, in press). The off-line sentence comprehension difficulties of some SLI children appear to be related to their more-limited functional working memory capacity and specifically to a difficulty coordinating the dual functions of memory storage and processing and managing the additional information processing demands posed by a conventional sentence comprehension task. By contrast, these children’s slower real-time sentence processing does not appear to be associated with their working memory limitation. Instead, it appears to be related primarily to inefficient lexical retrieval operations. Future research with SLI children and their normally developing peers might employ a wider range of sentence structures, larger language units, and on-line processing tasks to assess the potential relation between working memory and real-time language processing. Results from such studies would provide us with greater insights into how children with SLI manage their cognitive resources in the service of a range of language-related processing activities, thereby better informing our efforts to design and provide more sensitive treatments.

APPENDIX 1

<p>| Stimulus words organized by semantic category and word-referent size from smallest (top) to biggest (bottom) |
|--------------------------------------------------------|--------------------------------------------------------|</p>
<table>
<thead>
<tr>
<th>Clothing</th>
<th>Body</th>
<th>Parts</th>
<th>Transportation</th>
<th>Animals</th>
<th>Plant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socks</td>
<td>Eye</td>
<td>Skates</td>
<td>Mouse</td>
<td>Seed</td>
<td></td>
</tr>
<tr>
<td>Shoes</td>
<td>Thumb</td>
<td>Bike</td>
<td>Chick</td>
<td>Nut</td>
<td></td>
</tr>
<tr>
<td>Boots</td>
<td>Ear</td>
<td>Car</td>
<td>Cat</td>
<td>Leaf</td>
<td></td>
</tr>
<tr>
<td>Shirt</td>
<td>Foot</td>
<td>Truck</td>
<td>Dog</td>
<td>Bush</td>
<td></td>
</tr>
<tr>
<td>Coat</td>
<td>Head</td>
<td>Plane</td>
<td>Cow</td>
<td>Tree</td>
<td></td>
</tr>
</tbody>
</table>

APPENDIX 2

PILOT TESTING PROCEDURES AND RESULTS

Pilot testing was conducted in four phases. In the first phase, a stimulus sheet was created on which the five semantic categories were provided at the top of the page in separate columns. A set of 25 experimental stimulus words (5 words presented in 5 columns) was listed at the bottom of the page in random order. Twenty graduate students were asked to place each word from the bottom of the page in its respective semantic category. Then they were asked to arrange each word within each semantic category by size, beginning with the “smallest thing” (1) and ending with the “largest thing” (5). Subjects were instructed to think of “typical” or “general” instances of each word referent when they ordered the words by physical size. Results showed 100% agreement in semantic categorization. Within each semantic category, ordering by size ranged from 100% agreement
Montgomery: Working memory and sentence processing

(Transportation, Animals) to 93% agreement (Plant Things, Body Parts, Clothing). To derive 100% agreement in size ordering within these three categories, a second pilot phase (identical to the first) was completed. In this phase, those words in each category that yielded a different size order relative to the “majority” of respondents were excluded. Replacement words (i.e., highly familiar, monosyllabic words) from each of these three semantic categories were selected. As in the first phase, a different group of 20 graduate students was asked to categorize and order by size each of the words. Results showed 100% agreement in both semantic categorization and size ordering within each category. The third pilot phase entailed presenting another group of 20 graduate students with the actual dual-load experimental task in written form. For each word list, subjects were asked to order the words by (a) semantic category and (b) size within each group of semantically categorized words. Results showed 100% agreement in semantic categorization and size ordering. The fourth phase entailed presenting seven normally developing children (6;11 to 8;4) a set of 25 randomly ordered pictures (five pictures from each category) corresponding to the stimulus words. They were asked to make up five different groups/stacks of pictures, placing those pictures that go together in some way in the same group/stack first. Once they established their groupings, they were asked to order the pictures in each category by size, “starting with the smallest thing” and “ending with the biggest thing.” Results yielded 100% agreement in semantic categorization and 100% agreement in size ordering for Transportation and Body Parts. For Clothing, two children placed shoes before socks; for Animals, one child placed chick before mouse; and in Plant Things, two children placed nut before seed. Hence, when scoring the size ordering of words in the experimental proper, these variant orders were also scored as correct.

ACKNOWLEDGMENTS
This study was supported in part by a research grant (R29 DC 02535) from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. We express our gratitude to the children and their parents who participated in this study.

REFERENCES
Brown, L., Sherbenou, R., & Johnsen, S. (1990). Test of Nonverbal Intelligence (2nd ed.). Austin, TX: Pro-Ed.


