Differences in the Nonword Repetition Performance of Children With and Without Specific Language Impairment: A Meta-Analysis

Katharine Graf Estes
University of Wisconsin—Madison

Julia L. Evans
San Diego State University, San Diego, CA

Nicole M. Else-Quest
University of Wisconsin—Madison

Purpose: This study presents a meta-analysis of the difference in nonword repetition performance between children with and without specific language impairment (SLI). The authors investigated variability in the effect sizes (i.e., the magnitude of the difference between children with and without SLI) across studies and its relation to several factors: type of nonword repetition task, age of SLI sample, and nonword length.
Method: The authors searched computerized databases and reference sections and requested unpublished data to find reports of nonword repetition tasks comparing children with and without SLI.
Results: Children with SLI exhibited very large impairments in nonword repetition, performing an average (across 23 studies) of 1.27 standard deviations below children without SLI. A moderator analysis revealed that different versions of the nonword repetition task yielded significantly different effect sizes, indicating that the measures are not interchangeable. The second moderator analysis found no association between effect size and the age of children with SLI. Finally, an exploratory meta-analysis found that children with SLI displayed difficulty repeating even short nonwords, with greater difficulty for long nonwords.
Conclusions: These findings have potential to affect how nonword repetition tasks are used and interpreted, and suggest several directions for future research.

KEY WORDS: memory, meta-analysis, selective language impairment

The purpose of this study is to provide a quantitative summary of the difference in nonword repetition performance between children with and without specific language impairment (SLI), using a meta-analysis. Children with SLI display deficits in language acquisition and use despite having normal hearing and nonverbal intelligence, as well as no history of frank neurological impairments or emotional disturbances (Leonard, 1998). In addition to impairments in lexical, grammatical, and morphological development, children with SLI show substantial difficulty repeating nonsense words such as *loddernapish* compared with children with normal language skills (Bishop, North, & Donlan, 1996; Edwards & Lahey, 1998; Ellis Weismer et al., 2000; Gathercole & Baddeley, 1990; Montgomery, 1995b). Nonword repetition tasks have received a great deal of attention in investigations of language impairments because of their potential to contribute to understanding the underlying deficits of children with SLI and their potential as an identifier of children with SLI.

Although the finding that children with SLI repeat nonwords significantly less accurately than children with normal language (NL) occurs reliably across the literature, our meta-analysis indicates that the magnitude of the nonword repetition deficit for children with SLI exhibits a broad range across studies. We examine possible sources of this variability. The meta-analytic investigation of this growing literature has potential to inform how researchers understand the skills underlying nonword repetition performance, as well as the tasks used to measure these skills.

From the earliest studies of the nonword repetition performance of children with SLI, the skills tapped in the task were proposed to play an important role in language impairments. Kamhi and colleagues (Kamhi & Catts, 1986; Kamhi, Catts, Mauer, Apel, & Gentry, 1988) applied nonword repetition as a measure of phonological encoding because the task requires appropriate phonological perception and representation and minimizes memory storage and retrieval demands. On several phonological processing tasks, children with SLI and children with reading impairments showed similar patterns of difficulty. However, children with SLI performed significantly worse when repeating multisyllabic nonwords.

Gathercole and Baddeley (1989, 1990; reviewed in Baddeley, Gathercole, & Papagno, 1998) proposed that nonword repetition is a measure of *phonological working memory* (also referred to as the *phonological loop,* and more recently as *phonological short-term memory,* e.g., Gathercole, Tiffany, Briscoe, Thorn, & ALSPAC Team, 2005, or *verbal short-term memory,* Archibald & Gathercole, 2006), the component of memory that holds a temporary store of phonological information. This component is thought to be essential to the process of forming stable phonological representations to add to the lexicon and, therefore, to play a crucial role in vocabulary development. On the basis of longitudinal studies of nonword repetition performance and vocabulary development, Gathercole and colleagues proposed that early in language learning, phonological working memory skills support vocabulary growth (Gathercole & Baddeley, 1989; Gathercole, Willis, Emslie, & Baddeley, 1992). Conversely, after approximately age 5, vocabulary knowledge may support nonword repetition performance, such that children are increasingly able to apply lexical knowledge to repeat novel words. Thus, children with large vocabularies tend to perform well in nonword repetition tasks, although the strength of the relation with vocabulary diminishes with age (Gathercole & Adams, 1993; Gathercole et al., 1992). However, there is also evidence that lexical knowledge affects the nonword repetition performance of even younger children (Zamuner, Gerken, & Hammond, 2004) and that phonological working memory affects word learning through adolescence (Gathercole, Service, Hitch, Adams, & Martin, 1999).

In their study of children with SLI, Gathercole and Baddeley (1990) tested the prediction that if nonword repetition is an index of a skill that is important for language development (i.e., phonological working memory), then children with poor nonword repetition performance should have small vocabularies. They found that children with SLI (with small vocabularies) repeated significantly fewer nonwords correctly compared with children with normal language skills of a similar age matched on nonverbal intelligence, as well as a younger group matched on vocabulary, reading, and nonverbal measures.

There has been considerable debate surrounding the nature of the skills tapped in nonword repetition, whether it recruits phonological working memory (Bishop et al., 1996; Botting & Conti-Ramsden, 2001; Montgomery, 1995b; Van der Lely & Howard, 1993), phonological encoding (Kamhi & Catts, 1986), phonological awareness or sensitivity (e.g., Metsala, 1999), or a general phonological processing ability (e.g., Bowey, 1996, 2001). Many authors have also acknowledged that the act of repeating nonwords involves multiple processes (e.g., Briscoe, Bishop, & Norbury, 2001; Edwards & Lahey, 1998; Gathercole, Willis, Baddeley, & Emslie, 1994; Snowling, Chiat, & Hulme, 1991). A child's ability to repeat a novel word may be affected by any of the component skills involved in the process of hearing, encoding, and producing a word form: the ability to perceive speech distinctions; the preciseness, robustness, or organization of phonological and morphological representations; the ability to store the word form; and motor planning and articulation skills. The impairments of children with SLI may affect performance at any point or at many points in this process.

There is little disagreement that children with SLI show substantial difficulty in nonword repetition tasks. This difficulty has played a large role in the development of theories of SLI, including accounts of impairments based on deficits in phonological working memory capacity (Gathercole & Baddeley, 1990; Montgomery, 2002, 2003), phonological sensitivity or encoding (Kamhi & Catts, 1986; Metsala, 1999), and general limited processing capacity (Ellis Weismer et al., 2000; Marton & Schwartz, 2003). The degree of their difficulty in nonword repetition tasks has led to investigations of nonword repetition as a possible marker of SLI that can be used to identify impaired children (Bishop et al., 1996; Conti-Ramsden, Botting, & Faragher, 2001; Conti-Ramsden & Hesketh, 2003; Dollaghan & Campbell, 1998). Ellis Weismer and colleagues (2000) used the Nonword Repetition Test (NRT; Dollaghan & Campbell, 1998) and found that nonword repetition performance was a less culturally biased measure of language skills than typical standardized tests used to identify children with SLI (see also Campbell, Dollaghan, Needleman, & Janosky, 1997; Rodekohr & Haynes, 2001; Washington & Craig, 2004). Although the task does not perfectly identify SLI and NL groups, and it works best in conjunction with other language measures, the studies indicate that nonword repetition holds promise as a practical means of identifying children with SLI.

Further support for the application of nonword repetition as an identifier has come from Bishop and colleagues (1996), who found that children with persistent as well as resolved language impairments showed significant deficits in nonword repetition. Their twin study further revealed significant heritability of nonword repetition using the Children's Test of Nonword Repetition (CNRep; Gathercole et al., 1994) and led to the proposal that performance on the CNRep may act as a phenotypic marker of genetically based language impairments. The SLI Consortium (2002, 2004) used the CNRep in a genomewide quantitative trait locus analysis of SLI. They found a significant chromosome linkage for nonword repetition performance in individuals with SLI. Thus, nonword repetition is beginning to be applied broadly, from studies intended to identify children with SLI to studies intended to identify genes linked to language impairments.

Meta-Analysis Approach to Nonword Repetition

Nonword repetition tasks have received a great deal of attention because of their potential role in exposing the underlying impairments of SLI and because of their application in identifying children with SLI. Meta-analytical techniques can extend the capabilities of traditional reviews (e.g., Baddeley, 2003; Coady & Evans, in press; Montgomery, 2002, 2003) by presenting a quantitative summary of the body of findings, as well as by investigating whether there is variation in performance across studies that could be masked by reports of statistical significance. As is evident in the present analysis, nearly all studies comparing nonword repetition in children with and without SLI report statistically significant differences in performance, yet there remains substantial variability in the magnitude of the differences found. A further benefit of the meta-analysis approach is the power it provides. Bolstering statistical power in analyses is especially important in an area of research like language impairments, in which there are a relatively small number of studies addressing any particular issue and studies typically use relatively small numbers of participants.

This meta-analysis is based on an integration of effect sizes that highlights the magnitude of the difference in nonword repetition performance between children with and without SLI (Hedges & Becker, 1986). The results of each study are transformed into an effect size (d), which is the difference between the scores of children with NL and children with SLI, divided by the pooled standard deviation for both groups. Effect sizes can be interpreted in terms of standard deviation units. An effect size of $d = 1.0$ indicates that children with NL performed one standard deviation better than children with SLI. The calculation of d provides a common metric in order to combine findings across studies to find the overall mean difference in performance between groups (Durlak, 2003; see also Robey & Dalebout, 1998). The common metric also allows the comparison of effect sizes across studies using a variety of test designs and groups of children. This facilitates the identification of patterns across the literature in order to explore why children with SLI show different degrees of nonword repetition deficits.

The meta-analysis first provides a broad view of the literature, addressing the question: How large is the nonword repetition deficit of children with SLI? While this view clearly indicates that children with SLI display deficits, there remains a great deal that is not well understood. This meta-analysis addresses the following three questions about nonword repetition:

1. *Are all versions of the nonword repetition task interchangeable?* The versions of the task that have been used in research (i.e., CNRep, NRT) differ in potentially important design characteristics. Because the findings across nonword repetition measures are typically interpreted together in literature reviews, it is important to know whether the various versions all find a deficit of a similar magnitude.
2. *Are there age differences in the magnitude of the nonword repetition deficit in children with SLI?* While the clinical profile of children with SLI changes substantially over development, we know little about changes in basic processing abilities, such as those tapped in nonword repetition. This analysis will examine whether the magnitude of the SLI nonword repetition deficit is related to age across studies.
3. *Are there circumstances under which the nonword repetition skills of children with SLI are comparable to those of children with NL?* There are inconsistent findings regarding whether children with SLI are impaired at repeating nonwords of all lengths (one to five syllables) or only long nonwords. Resolving the ambiguity of these findings promises to inform understanding of phonological memory capacity in children with SLI.

The first and second questions are addressed in *moderator analyses,* a type of analysis performed when there is significant variability in the effect sizes in a meta-analysis. It provides a means to examine causes of the variability (i.e., factors that moderate the variability) by testing factors that are hypothesized to affect performance across studies (Robey & Dalebout, 1998). Thus, the broad analysis of effect sizes is followed by more fine-grained analyses investigating what factors affect the size of the nonword repetition deficit. In a domain in which there is a general consensus that children with SLI are significantly impaired, the greatest contribution of the meta-analysis may come from the deeper understanding that the moderator analyses provide about this popular and potentially revealing task. Given the relatively small number of studies in the analysis, to prevent a statistical "fishing expedition," we have limited the moderator analyses to two factors hypothesized to affect effect sizes: type of nonword repetition task and the age of the SLI sample tested (Robey & Dalebout, 1998).

We first address the question: Are all versions of the nonword repetition task interchangeable? Studies using different measures of nonword repetition performance (e.g., CNRep, Gathercole et al., 1994; NRT, Dollaghan & Campbell, 1998) are regularly grouped together in literature reviews, as authors summarize the findings of significant deficits in phonological memory or the task's utility as an identifier (e.g., Conti-Ramsden, 2003; Gray, 2003; Montgomery, 2002, 2003). On the surface, this seems sensible, because all of the tasks ostensibly measure the same skill. However, common measures of nonword repetition differ in important characteristics that could affect performance, such as wordlikeness, articulatory complexity, and nonword length (see Table 1). It is not surprising that the versions of the task differ; different researchers designed them, often with different motivations behind the design decisions (see Coady & Evans, in press). However, because of these design differences, it is not clear whether all measures of nonword repetition likely tap deficits of the same magnitude and character. Because of the theoretical and applied value of the nonword repetition task, it is essential to understand what we are measuring when using it. We address whether versions of the nonword repetition task are interchangeable using a moderator analysis. The analysis indicates whether the size of the nonword repetition deficit in children with SLI is consistent across measures.

Wordlikeness is one design characteristic that may impact effect size magnitude across studies. Children with NL and children with SLI both show superior nonword repetition performance for high-wordlike nonwords compared with low-wordlike nonwords (Coady, Evans, & Kluender, 2006; Gathercole, 1995; Munson, Kurtz, & Windsor, 2005). Furthermore, Gathercole found that performance on low-wordlike nonwords was linked to digit span, a conventional measure of phonological memory. Gathercole argued that repetition of low-wordlike nonwords was a more direct measure of phonological working memory because it reduces the opportunity to rely on representations of known words, whereas repetition of high-wordlike nonwords receives support from extant lexical knowledge.

One commonly used measure of nonword repetition, the CNRep, includes items of both high and low wordlikeness (see Gathercole, 1995; Gathercole et al., 1994); however, both item sets include embedded English words and affixes (e.g., *perplisteronk;* see also Dollaghan, Biber, & Campbell, 1995, for facilitative effects of embedded words). In contrast, another measure, the NRT (Dollaghan & Campbell, 1998), was designed to minimize wordlikeness by excluding embedded real words and assigning consonants to rare positions.

High-wordlike test items may lead to an increased advantage for children with NL who are able to take advantage of their lexical knowledge. Children with SLI, because of their smaller vocabularies, degraded lexical representations (Bishop, 2000), or differently organized lexicons (Dollaghan, 1998), may not have lexical support readily available for repeating nonwords, in addition to whatever deficit exists in basic phonological memory. Thus, high-wordlike nonwords could produce larger differences between children with and without SLI than low-wordlike nonwords. Alternatively, when nonword repetition tasks consist of low-wordlike items, children with SLI may fall most behind their peers because there is no opportunity to support their already fragile phonological processing with lexical knowledge (see Munson et al., 2005). There is mixed evidence regarding the effects of wordlikeness in children with SLI (Briscoe et al., 2001; Coady et al., 2006; Munson et al., 2005); therefore, the direction of its effect on the nonword repetition deficit is unclear. In any case, there are indications that nonword repetition tasks differing in the wordlikeness of their test items (i.e., CNRep and NRT) might recruit different underlying influences on performance, which may in turn affect the magnitude of the difference in performance between children with and without SLI.

Nonword repetition tasks also differ in their level of articulatory complexity (see Table 1). Although children with SLI do not typically have clinical articulation problems (see Leonard, 1998), they have more difficulty repeating complex nonwords (Bishop et al., 1996; Briscoe et al., 2001; but see Edwards & Lahey, 1998). Stark and Blackwell (1997) found evidence of subtle oral motor deficits in children with SLI (see also Goffman, 2004); such deficits could compromise performance on complex test items in particular. The CNRep contains both "complex" (containing consonant clusters) and "simple" (only singleton consonants) nonwords; the nonword sets designed by Montgomery (e.g., Montgomery, 1995b, 2004) also contain some clusters. The NRT was designed to minimize complexity by excluding clusters and late-acquired consonants. Thus, the more articulatorily complex nonword repetition tasks may tap subtle subclinical articulation or motor planning difficulties in children with SLI, which could promote larger effect sizes.

An additional difference across tasks is the scoring method used. If children with SLI are more likely than children with NL to make slight errors, then an

Table 1. Characteristics of nonword repetition measures.

Measure	No. of items	No. of syllables	Wordlikeness	Articulatory complexity	Scoring method
CNRep					
Gathercole et al. (1994)	40[a]	2–5 (Gathercole & Baddeley, 1990, used 1–4 syllables)	Mixed (see Gathercole, 1995) Contain English words and affixes	Mixed 20 "simple"with singleton consonants only, 20 "complex" with consonant clusters	Total number correct Credit only for words with no errors other than consistent phoneme substitutions
NRT					
Dollaghan & Campbell (1998)	16	1–4	Low No English words, consonants in infrequent positions	Low No late acquired consonants,[b] no consonant clusters	Percentage of phonemes correct[c]
Montgomery set					
Montgomery (1995b)	40 or 48[d]	1–4	Mixed Produced with variable stress, contain English words and affixes, wordlikeness rating range available in Montgomery (1995a)	Mixed Some with late acquired consonants, include consonant clusters	Total number correct
Three- to four-syllable items					
Coady et al. (2006)[e]	24	3–4	Mixed 12 high-phoneme frequency, 12 low-phoneme frequency	Mixed/low No consonant clusters, include late acquired consonants	Percentage of phonemes correct
Edwards & Lahey (1998)	6	3–4	Mixed/low One syllable was an English word, wordlikeness ratings available	Mixed Some with consonant clusters, include late acquired consonants	Percentage of phonemes correct
Munson et al. (2005)	20	3–4	Mixed 10 high-phonotactic probability/wordlikeness, 10 low	Mixed/low No consonant clusters, include late acquired consonants	Percentage of phonemes correct
Kamhi & Catts (1986)	30	3–4	Mixed/high Derived from English words, contain English words and affixes	Mixed 15 simple with "clear phonemic contrasts," 15 complex with "minimal phonemic contrasts"	Total number correct
Kamhi et al. (1988)	40	3–4	Mixed/high Derived from English words, contain English words and affixes	Mixed Some with "clear phonemic contrasts," some with "minimal phonemic contrasts"	Total number correct

Note. CNRep = Children's Test of Nonword Repetition; NRT = Nonword Repetition Test.

[a]Gray (2003) used 20 items from the CNRep in initial testing; values in meta-analysis are based on the first test of the 20 items. [b]Late acquired consonants based on "Late Eight" consonants (Shriberg & Kwiatkowski, 1994). But see scoring in Gray (2004). [d]Horohov and Oetting (2004) used 20 of the items in Montgomery (1995b). [e]Coady et al. (2006) values based on analysis of the phoneme frequency manipulation items only.

all-or-nothing scoring method (typically used with the CNRep), in which children get credit only for completely accurate repetitions, may place children with SLI at a greater disadvantage than when they can receive partial credit (i.e., the percentage-of-phonemes-correct scoring method that is used with the NRT).

Another important difference across nonword repetition measures is the length of the nonwords, specifically the number of syllables used in the test items (a topic that is addressed again later). Children with SLI show consistent deficits repeating long nonwords (with three or more syllables) but show smaller or nonsignificant difficulty repeating shorter nonwords (Bishop et al., 1996; Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990; Montgomery, 1995b). Thus, a nonword repetition measure that consists primarily of long nonwords should be expected to produce large effect sizes. The CNRep contains two- to five-syllable items; the NRT and the nonword sets used by Montgomery and colleagues contain one- to four-syllable items. There are also several studies that use only three- and four-syllable nonwords (Coady et al., 2006; Edwards & Lahey, 1998; Kamhi & Catts, 1986; Kamhi et al., 1988; Munson et al., 2005). Based on nonword length, measures like the CNRep that contain many long items may promote larger differences between children with and without SLI than measures with many short items.

The difference in nonword length across measures raises the question of whether all nonword repetition tasks tax children with SLI equally. A final design factor that may affect the difficulty of the task is the number of test items. The measures range from 6 (Edwards & Lahey, 1998) to 48 nonwords (Montgomery, 1995b). If the already vulnerable language processing skills of children with SLI are taxed, such as in a long task that leads to fatigue, their performance may begin to fall farther behind their peers as the task continues.

Taken together, the preceding literature review indicates that it is likely that all versions of the nonword repetition task are not interchangeable because of differences in design. There is experimental evidence to this effect. Recently, Archibald and Gathercole (2006) examined the performance of children with and without SLI on both the CNRep and the NRT. The authors reported that there was evidence for a larger nonword repetition deficit on the CNRep: In analyses adjusted for differences in nonverbal ability, children with SLI performed less accurately than both age-matched and language-matched control groups on the CNRep. In contrast, children with SLI did not perform significantly less accurately than the two control groups on the NRT. (However, a comparison of only the SLI and age-matched groups found a significant difference in performance, suggesting a lack of power for the analysis of the three groups together.) Archibald and Gathercole's investigation of task characteristics revealed that long nonwords produced a larger difference between groups than short nonwords, supporting the role of impaired verbal short-term memory in children with SLI. However, the authors also suggested that the difficulty of children with SLI on the relatively wordlike CNRep items (compared with the low-wordlike NRT items) could indicate that a lack of support from lexical representations also plays a role in the nonword repetition deficit. In addition, the performance of children with SLI on the CNRep was more impaired for articulatorily complex items than for simple ones. Archibald and Gathercole concluded that the CNRep may yield greater differences between children with and without SLI because of a convergence of factors that make nonword repetition difficult for children with SLI: lexical mediation, articulatory complexity, and long nonwords.

The present meta-analysis expands on Archibald and Gathercole's (2006) experimental findings by examining performance across many samples as well as additional nonword repetition measures. We compare the average effect size for studies using four types of nonword repetition measures: (a) CNRep (Gathercole et al., 1994); (b) NRT (Dollaghan & Campbell, 1998); (c) lists using three- to four-syllable words (e.g., Edwards & Lahey, 1998; Kamhi & Catts, 1986); and (d) nonword sets designed by Montgomery and colleagues (e.g., Montgomery, 1995b, 2004). If the measures yield dissimilar effect sizes for the nonword repetition deficit in SLI, it will suggest that design characteristics of the measures affect performance and indicate that the measures are not interchangeable gauges of the same deficit. Future experiments will then be necessary to clarify the origins of the differences across measures, as it is not within the scope of this meta-analysis to selectively examine the influence of individual task characteristics. As demonstrated in Table 1, these factors are generally confounded in the measures. Alternatively, despite the diversity in nonword repetition tasks, they may all yield similar magnitude effect sizes. This finding would suggest that despite differences in the characteristics of the measures such as wordlikeness, complexity, and nonword length, the magnitude of the difference between children with and without SLI is robust.

Our second question is also addressed in a moderator analysis: Are there age differences in the magnitude of the nonword repetition deficit in children with SLI? While studies of children with NL have traced the developmental trajectory of nonword repetition performance (e.g., Gathercole et al., 1992; reviewed in Baddeley et al., 1998), no previous research has investigated the change over time in children with language disorders. Bishop (1992) pointed out that developmental change in children with SLI is a factor that is commonly overlooked. Similarly, Thomas and Karmiloff-Smith (2005) have argued that in order to understand developmental disorders, it is essential to take into account the developmental process; they are not static systems. In addition, understanding whether there are age differences in the nonword repetition deficit of children with SLI is important for the use of the task as an identifier of language impairments. It is not clear if there are age differences in the task's utility. Each of the studies in the analysis compared children with SLI with a control group, primarily of the same age. This analysis examines whether there is change in the distance between children with SLI and children with NL across the ages tested.

There are several possible forms that age differences in the nonword repetition deficit could take. First, the performance of children with SLI may be the most different from their peers at a young age. Gathercole (1995) proposed that the influence of vocabulary size on nonword repetition performance plateaus as children gain sufficiently large vocabularies to support performance. With vocabulary growth, children with SLI may gain access to a critical threshold of lexical support, which they could use to supplement any deficits in the underlying memory skills also tapped in the task. Thus, although the nonword repetition skills of children with SLI may be delayed, they may eventually catch up to children with NL.

Alternatively, children with SLI may fall more behind their typically developing peers with age. In children with NL, nonword repetition is initially supported by phonological working memory that is later supplemented by lexical knowledge, and their performance improves substantially with age (Gathercole et al., 1994). Children with SLI may initially be at a disadvantage because of poor phonological working memory skills. This disadvantage could be compounded by difficulty recruiting lexical knowledge. As lexical support becomes available to children with NL, but not to children with SLI, the difference in their performance may grow with age. Also, children who retain the SLI diagnosis at older ages may make up a more focused, more impaired sample, as they represent the children whose clinically significant language problems have not been resolved by adolescence (Bishop, 1992).

Finally, a third hypothesis is that the nonword repetition deficit of children with SLI may remain consistent across development. The performance of children with SLI and their peers with NL may improve at a similar pace. While the language profile of children with SLI has been shown to change over development (see Leonard, 1998), nonword repetition may tap skills that remain consistently impaired. Bishop and colleagues' (1996) study of children with resolved language impairments provides some evidence for such consistency. They found that children who no longer met criteria for a language disorder remained impaired on measures of nonword repetition; their performance was similar to that of children with persistent language impairments.

A longitudinal study specifically examining sensitive measures of developmental level in addition to age would be best suited for examining developmental changes in nonword repetition performance. However, the meta-analysis provides a first step by allowing us to take advantage of the existing data and combine the findings from throughout the literature. Across studies, the children with SLI ranged in average age from 4 to 12 years of age. In a moderator analysis, we investigate whether variability in the effect sizes across nonword repetition studies is related to the age of the children with SLI tested.

The third question we investigate is the following: Are there circumstances under which the nonword repetition skills of children with SLI are comparable to those of children with NL? To address this question, we examine the performance of children with SLI on nonwords of different lengths. Children with SLI have consistently displayed significant deficits when repeating long nonwords of three or more syllables (Bishop et al., 1996; Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990; Montgomery, 1995b). This length effect has been demonstrated in two ways. Children with SLI have shown impaired performance for long nonwords compared with (a) their NL (age-matched and occasionally language-matched) peers and (b) their own performance on short nonwords (one to two syllables).

The poor performance of children with SLI at repeating long nonwords has been used to support hypotheses about the underlying causes of nonword repetition deficits and the language impairment more generally. Gathercole and Baddeley (1990) argued that if the repetition deficit is due to impaired storage in the phonological loop component of working memory, children with SLI should have greater difficulty repeating long sequences (see also Montgomery, 2004). While children with NL also repeat long nonwords less accurately than short nonwords, the reduced phonological memory resources of children with SLI may be more readily overwhelmed by long nonwords and performance may break down to a greater extent than for children with NL. Furthermore, it has been proposed that the nonword repetition performance of children with SLI represents a general lack of phonological working memory capacity that is at the root of other language difficulties, such as the ability to comprehend long sentences (e.g., Montgomery, 1995b).

There have been mixed findings regarding whether children with SLI perform below their peers at repeating short nonwords (Ellis Weismer et al., 2000), with several studies reporting significant deficits for long nonwords only (Bishop et al., 1996; Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990; Montgomery, 1995b). Thus, a common conclusion is that children with SLI do not have difficulty repeating short nonwords (e.g., Coady & Evans, in press; Montgomery, 2002, 2003)—items that may be within the memory limitations of children with SLI. Interestingly, one study that found a significant impairment in children with SLI on short nonwords was the study with the largest sample size (Ellis Weismer et al., 2000). This finding raises the possibility that the nonsignificant differences between children with and without SLI on short nonwords have been influenced by a lack of statistical power in the small samples typically analyzed.

It is unclear how evidence of substantial difficulty with short nonwords fits within the limited phonological working memory capacity account. While children with SLI likely exhibit greater deficits for longer nonwords, poor performance repeating even one-syllable items may implicate factors in addition to limited phonological working memory capacity, such as impaired phonological encoding, degraded phonological representations, or poor lexical support. Presumably, a single novel syllable should not be sufficient to overwhelm the capacity of the working memory system. Or if it is sufficient, this phenomenon deserves explanation as well.

In a second exploratory meta-analysis, we examine effect sizes for studies that reported nonword repetition performance broken down by nonword length. Meta-analysis techniques afford the combination of samples across studies, providing a larger sample than any individual study has examined. With this large sample we can, with some confidence, investigate whether children with SLI are impaired at repeating short nonwords. This analysis also examines changes in effect sizes as nonwords add syllables. The findings will shed light on the role of limitations in phonological working memory capacity and other factors that could affect performance in the seemingly simple task of repeating one- and two-syllable sequences.

In summary, we predict that there will be a very large difference in the performance of children with and without SLI on tests of nonword repetition. This prediction is to be expected, as nearly all studies of nonword repetition report significant differences between children with and without language impairments. A novel contribution of the present meta-analysis is that it provides a quantitative summary of the magnitude of the SLI deficit in nonword repetition. Because this deficit is relatively well established, the average effect size value will provide a useful benchmark for future research, as meta-analysis is applied to understand the impairments of children with SLI about which there is less consensus (e.g., motor and nonverbal cognition impairments).

The first moderator analysis investigates whether the type of nonword repetition measure used affects the magnitude of difference between children with and without SLI. The analysis will inform understanding of the nature and magnitude of the nonword repetition deficit. The second moderator analysis explores age differences in this deficit that is thought to be fundamental to SLI. Finally, in a set of exploratory analyses, we examine the magnitude of the difference between children with and without SLI at several nonword lengths, an analysis that has potential to affect accounts of the underlying deficits of SLI.

Method

Identification of Studies

Computerized database searches of PsychINFO and Web of Knowledge were conducted to identify articles reporting nonword repetition performance in children with and without SLI. Because nonword repetition measures were not always clearly identified in abstracts, we searched using the broad terms *language impairment and memory, SLI and memory, SLI and nonword, language impairment and nonword,* and *language disorder and nonword.* Abstracts were collected and saved if studies reported data for children with a specific or primary language impairment (generally referred to as SLI, but not all authors used this term; e.g., Bishop et al., 1996; Munson et al., 2005; Washington & Craig, 2004) and a typically developing control group. The saved abstracts also mentioned a task that could refer to a nonword repetition measure (i.e., phonological [working] memory, verbal memory, nonsense words). The search yielded 44 saved articles from PsychINFO and 45 from Web of Knowledge, with many appearing in both databases. In addition, we viewed articles discussed in a recent review of nonword repetition in SLI by Coady and Evans (in press) as well as articles listed in the references of gathered articles.

A common concern in meta-analyses is the "file drawer problem" (Robey & Dalebout, 1998; Rosenthal, 1979), which is the tendency for statistically significant findings to be published more easily and more often than nonsignificant findings. This can lead to inflated overall effect size estimates. However, the publication bias against nonsignificant findings is likely to be less of a problem in the study of children with language impairments than many other domains, because cases in which children with SLI show performance similar to that of their peers are also of interest. To avoid the effects of publication bias, we sought to include unpublished data by contacting authors who have previously performed studies of nonword repetition in SLI to request unreported data. Four authors had data available to contribute to the analysis (Archibald & Gathercole, 2006; Coady et al., 2006; see also Coady, Evans, Kluender, Mainela-Arnold, & Ryan, 2003; Montgomery & Windsor, 2004; Munson et al., 2005). We also included an unpublished data set from our own laboratory (Evans, Graf Estes, Coady, Ryan, & Simon, 2005). In summary, the computerized search, reference search, and data requests yielded 60 studies considered for inclusion.

Inclusion Criteria

Studies were included in the analysis if they presented original data comparing children with SLI to typically developing children matched on age, language skill, or nonverbal intelligence. The language-impaired group met broad criteria for SLI: impaired expressive and/or receptive language skills and normal nonverbal intelligence. Most studies also reported that children met further criteria such as normal hearing and no known history of frank neurological impairments or social–emotional disorders. The studies used a variety of criteria for defining the language-impaired group (e.g., diagnosis by a speech-language pathologist; poor language skills and normal nonverbal intelligence, as measured by various standardized tests).

We included studies using nonword repetition tasks that required children to listen to and repeat, one at a time, a series of nonsense words. Tasks that required children to repeat an entire list of nonwords from memory were excluded from the analysis because they may tap different aspects of memory than tasks that require individual repetitions.

Studies were excluded from the meta-analysis if participants were selected to represent extreme nonword repetition scores (e.g., Botting & Conti-Ramsden, 2001) or if participants had a history of language impairments but current status was unknown or the language impairment had resolved (e.g., Bishop et al., 1999). When it was possible to determine that nonword repetition performance was reported more than once for the same children or a subgroup of children, the study with the larger sample size was used in the analysis unless other inclusion or exclusion criteria applied (e.g., Montgomery, 1995a, 1995b). Thus, only independent samples were included in the analysis. Studies were also excluded if the data available were insufficient to calculate effect sizes (e.g., missing means and standard deviations or test statistics). One study (Marton & Schwartz, 2003) was excluded because no other study used the same nonword repetition measure; therefore, it could not be included in the moderator analysis. The data from Marton and Schwartz were included in the exploratory nonword length analysis.

Coding. For each study, we coded the following information: (a) all statistics regarding differences in nonword repetition performance between SLI and control groups, including means, standard deviations, *t* tests, and *F* tests; (b) number of SLI and control participants; (c) mean age of participants; (d) SLI diagnosis criteria; (e) type of control group(s) used (language matched [LM], chronological age matched [CAM], or nonverbal IQ matched); (f) nonword set used (CNRep, NRT, Montgomery nonword set, and three- to four-syllable set); (g) number of nonwords; (h) whether the nonwords exhibited (English) wordlike characteristics, as described by the original authors; (i) articulatory complexity, as described by the original authors; and (j) number of syllables of nonwords.

Of the 23 studies meeting the inclusion criteria, a random sample of 5 (22%) of the studies was recoded.[1] Coder agreement was 99%, and disagreements were resolved through reexamination of the original articles and discussion.

Final sample of studies. Twenty-three studies from eight different journals were included in the meta-analysis. These studies reported 25 effect sizes across 549 children with SLI and 942 children with NL, as shown in Table 2. Three studies contributed two effect sizes each because two separate groups of children with SLI and two suitable NL control groups were tested (see Table 2). Several studies compared one group of children with SLI to both a CAM sample and a LM sample. Table 2 presents data from both control groups, when available. However, in keeping with the requirement of independent samples in a meta-analysis, only one effect size for each SLI sample was entered into the analysis. In these cases, we selected the effect size from the CAM sample in order to maintain a consistent control group (CAM samples = 23, LM samples = 2). The excluded effect sizes are shown with asterisks and in italics in the table.[2]

Analysis

We calculated an effect size, *d,* for each sample by subtracting each SLI group's mean nonword repetition score from its control group's mean score and dividing the difference by the mean within-group (pooled) standard deviation.

$$d = (M_{\text{CONTROL}} - M_{\text{SLI}})/SD_{\text{WITHIN GROUP}}$$

Thus, positive effect sizes indicate higher performance by control group children. According to Cohen's (1988) guidelines for the interpretation of effect sizes, d = .20, .50, and .80 are considered small, medium, and large, respectively. For some studies, effect sizes were derived from *t* tests or *F* values from analyses of variance (ANOVAs) using the procedure described in Hedges and Becker (1986).

All effect sizes were corrected for bias in estimation of population effect sizes using the formula provided in Hedges and Becker (1986). The correction factor takes into account the sample size associated with each effect size in order to reduce the impact of small samples on the overall effect size, because small samples are more likely

[1] The second coder did not recode Items g–i because the first coding revealed that they were identifiable from coding of the type of nonword repetition measure.

[2] The effect of control group is another interesting factor that likely contributes to differences in effect sizes across nonword repetition studies. However, because of the small number of studies and the imbalance in control group types across measures of nonword repetition (see Table 2), it was not feasible to investigate the difference in effect sizes for LM and CAM control group comparisons.

Table 2. Studies of nonword repetition differences in children with and without SLI (organized by nonword repetition measure).

Study	Nonword set	Control group type	SLI *n*	Control group *n*	Effect size[a] (*d*)	Age of SLI group	Age of control group
Archibald & Gathercole (2006)[b]	CNRep	CAM	12	12	**1.52**	9.67	9.75
Archibald & Gathercole (2006)	*CNRep*	*LM[c]*	*12*	*12*	*.37*	*9.67*	*6.08*
Bishop et al. (1996)	CNRep	CAM	39	79	**1.48**	7.83	7.70
Briscoe et al. (2001)							
Younger: CAM	CNRep	CAM	14	20	**3.00**	8.96	8.49
Briscoe et al. (2001)							
Younger: LM	*CNRep*	*LM[c]*	*14*	*15*	*1.90*	*8.96*	*7.40*
Conti-Ramsden (2003)	CNRep	CAM	32	32	**1.73**	5.08	4.75
Conti-Ramsden & Hesketh (2003)							
Language match	*CNRep*	*LM[d]*	*32*	*32*	*0.06*	*5.08*	*2.83*
Farmer (2000)							
Language unit: CAM	CNRep	CAM[c]	8	8	**3.19**	10.78	10.33
Farmer (2000)							
Special School: LM	CNRep	LM	8	8	**1.93**	11.04	8.65
Gathercole & Baddeley (1990)	CNRep	NVIQ	5	5	**4.34**	8.50	7.70
Gathercole & Baddeley (1990)							
Language match	*CNRep*	*LM[c,e]*	*5*	*5*	*5.45*	*8.50*	*6.50*
Gray (2003)	CNRep	CAM	22	22	**3.86**	5.01	5.00
Archibald & Gathercole (2006)	*NRT*	*CAM*	*12*	*12*	*1.78*	*9.67*	*9.75*
Archibald & Gathercole (2006)[b]	*NRT*	*LM[c]*	*12*	*12*	*−0.26*	*9.67*	*6.08*
Dollaghan & Campbell (1998)	NRT	CAM	20	20	**1.89**	7.80	7.80
Ellis Weismer et al. (2000)	NRT	CAM	80	359	**0.65**	7.90	8.00
Evans et al. (2005)							
Older	NRT	CAM	14	17	**1.46**	12.78	12.34
Evans et al. (2005)							
Younger	NRT	CAM	26	62	**1.05**	9.02	8.71
Gray (2004)	NRT	CAM	20	20	**2.01**	4.87	4.82
Rodekohr & Haynes (2001)							
African American	NRT	CAM	10	10	**0.87**	7.20	7.20
Rodekohr & Haynes (2001)							
White	NRT	CAM	10	10	**1.30**	7.20	7.20
Washington & Craig (2004)	NRT	CAM	25	56	**0.62**	4.64	4.64
Coady et al. (2006)	3–4 syllable	CAM	20	20	**1.05**	9.25	8.58
Edwards & Lahey (1998)	3–4 syllable	CAM/NVIQ	54	54	**0.85**	6.90	6.90
Kamhi & Catts (1986)	3–4 syllable	CAM/NVIQ	12	12	**2.72**	7.90	7.20
Kamhi et al. (1988)	3–4 syllable	CAM	10	10	**2.09**	8.00	7.60
Munson et al. (2005)							
Age match	3–4 syllable	CAM	16	16	**1.42**	11.25	11.17
Munson et al. (2005)							
Language match	*3–4 syllable*	*LM[c]*	*16*	*16*	*−0.45*	*11.25*	*7.50*
Horohov & Oetting (2004)							
Age match	Montgomery	CAM	18	18	**2.12**	6.23	5.98
Horohov & Oetting (2004)							
Language match	*Montgomery*	*LM[c]*	*18*	*17*	*1.36*	*6.23*	*4.47*
Montgomery (1995b)	Montgomery	LM[f]	14	13	**2.16**	8.20	6.80
Montgomery (2004)							
Age match	Montgomery	CAM	12	12	**1.18**	8.75	8.67
Montgomery (2004)							
Language match	*Montgomery*	*LM[f]*	*12*	*12*	*0.93*	*8.75*	*6.83*
Montgomery & Windsor (2004)	Montgomery	CAM	48	48	**1.91**	8.78	8.74

Note. Effect sizes in bold were included in the analysis, items in italics were not included in the analysis. CAM = chronological age-matched group; LM = language-matched group; NVIQ = nonverbal intelligence score–matched group; SLI = specific language impairment.

[a]Effect size (*d*) is the unweighted, uncorrected effect size. [b]To maintain independent samples in the analysis, we included the data from Archibald and Gathercole's (2006) comparison of the CAM and SLI groups on the CNRep only. LM control group identifiers are as follows: [c]matched on receptive vocabulary, [d]general receptive and expressive language, [e]reading, and [f]receptive syntax.

to produce extreme values (Robey & Dalebout, 1998). Each effect size was multiplied by the correction factor, which reduced the magnitude of effect sizes that came from small samples (e.g., Gathercole & Baddeley, 1990). The formula for the correction factor (CF) is

$$CF = (1 - \{3/[(4 \times N_{\mathrm{CONTROL}}) + (4 \times N_{\mathrm{SLI}}) - 9]\}),$$

where N_{CONTROL} and N_{SLI} refer to the sample size of the control and SLI groups, respectively. Analyses were conducted using syntax commands in SPSS that were based on the formulas described by Hedges and Becker (1986).

Fixed-effects model. We chose to analyze the data using a fixed-effects model in order to make conditional inferences based only on the samples used in the analysis, assuming that effect size parameters are fixed but unknown constants (Hedges & Vevea, 1998). Robey and Dalebout (1998) stated that a fixed-effects model is most appropriate when nearly all of the studies in a domain have been included in the analysis and the population of studies are similar in their design, goal, and outcome. The studies in the present analysis meet these criteria. We performed a thorough search for published and unpublished data, and all of the studies included a task in which children with and without SLI repeated nonsense words with the expectation that children with SLI would have more difficulty on the task (in particular, when compared with CAM controls).

Results

Magnitude of Overall Difference in Nonword Repetition Performance

The individual effect sizes and corresponding study characteristics are presented in Table 2. The weighted mean effect size across the 25 effect sizes was 1.27, indicating that children with SLI performed on average more than a full standard deviation below the children with NL on nonword repetition tasks. The 95% confidence interval (CI) was from 1.15 to 1.39. All of the effect sizes in the analysis were positive, indicating higher performance in the control groups than in the SLI groups. Homogeneity analysis, using the procedures described in Hedges and Becker (1986), revealed significant nonhomogeneity among the effect sizes, $H_T = 117.06$, $p < .05$. This indicates that there was statistically significant variability among the effect sizes.

Moderator Analyses

The homogeneity analysis tests whether all of the studies in the meta-analysis share a common effect size. The H_T statistic has a chi-square distribution with $k - 1$ degrees of freedom; when it is significant, it can be assumed that the sample of effect sizes does not share a common effect size (i.e., it is significantly nonhomogeneous), and moderator analyses should be conducted to determine sources of the nonhomogeneity (Hedges & Becker, 1986). In addition, H_T can be partitioned, as in the ANOVA model, such that $H_T = H_B + H_W$. Thus, when we follow up the H_T test with moderator analyses, we can calculate the H_B (analogous to the F statistic for between-groups differences and having a chi-square distribution of $p - 1$) to determine whether effect sizes vary according to the levels of the moderator. H_W reflects the remaining variability in effect sizes after partitioning out the H_B.

To explain the nonhomogeneity in effect sizes, we analyzed two potential moderators: type of nonword repetition task and age.

Nonword repetition task. We investigated whether variability in effect sizes could be attributed to the different measures used to test nonword repetition. Results for this analysis are shown in Table 3. The studies were divided based on the type of nonword repetition task used, including CNRep, NRT, three- to four-syllable nonword sets, and the Montgomery set. All of the effect sizes were statistically significant, large, and positive, indicating that children with NL performed better than children with SLI on all versions of the task. The CNRep set yielded the largest effect size, followed by the Montgomery set, the three- to four-syllable lists, and the NRT. Overall, the test of between-groups homogeneity was significant ($H_B = 50.69$, $p < .05$), indicating that the nonword repetition measure was significantly related to the magnitude of effect sizes. The within-group homogeneity test was also significant ($H_w = 66.57$, $p < .05$), indicating that there was a remaining significant nonhomogeneity among the effect sizes. That is, the between-groups and within-group measures of homogeneity show that, while the type of nonword repetition measure used explained a statistically significant amount of the variability in effect sizes (indicated by H_B), there remained significant variability that was not explained by this moderator (indicated by H_w).

To examine model fit, we ran three multiple regression analyses with nonword set predicting effect size magnitude, using the inverse of effect size variance as a case weight.

Table 3. Effect sizes as a function of nonword repetition measure.

Nonword repetition measure	k	d	95% CI	H_W
CNRep	8	1.94	1.66–2.21	29.20*
NRT	8	0.90	0.72–1.07	21.88*
Other 3- to 4-syllable item sets	5	1.16	0.87–1.44	12.29*
Montgomery set	4	1.83	1.49–2.18	3.20

Note. CI = confidence interval; k = the number of effect sizes; H_W = the within-group homogeneity statistic (Hedges & Becker, 1986).
*Significant nonhomogeneity, $p < .05$, chi-square test.

In the first analysis, we compared the CNRep (coded 1) with the NRT (−1). This contrast was significant (β = .60, z = 7.47, p < .05, CI = .44–.75), explaining 35.8% of the variance in effect size magnitude and indicating that the CNRep and NRT measures produced significantly different effect sizes. However, model fit was rejected (H_E = 74.24, p < .05), indicating that factors in addition to type of nonword repetition task contributed to the nonhomogeneity of the sample of effect sizes. In the second regression analysis, we compared the CNRep (1) with the three- to four-syllable sets (−1). The contrast was significant (β = .37, z = 3.63, p < .05, CI = .17–.57), explaining 13.5% of the variance in effect sizes and indicating that the CNRep and the three- to four-syllable lists produced significantly different effect sizes. Again, model fit was rejected (H_E = 101.40, p < .05), indicating that factors in addition to nonword set contributed to the nonhomogeneity. In the third regression analysis, the contrast compared the CNRep (1) with the Montgomery nonword set (−1); this contrast was not significant (β = .17, z = 1.49, p > .05, CI = −.06–.39), explaining only 2.7% of the variance in effect sizes and indicating that the CNRep and Montgomery nonword set did not produce significantly different effect sizes. Model fit was rejected (H_E = 114.09, p < .05), indicating that other factors were responsible for the nonhomogeneity of the sample of effect sizes.

Age. To determine whether effect sizes attenuated or increased with age (see Table 2), we ran a multiple regression analysis using age as the predictor variable and corrected effect size as the outcome variable; the inverse of the variance of each effect size served as the regression weight (Hedges & Becker, 1986). The regression model was not significant, indicating that effect size magnitude was not associated with age, β = .05, $F(1, 23) = 0.07$, p > .05.

Exploratory Meta-Analyses

In addition to the moderator analyses, we examined the effect sizes across nonword lengths (one-, two-, three- and four-syllable items) by using a subset of the original 23 studies reporting nonword repetition scores by nonword length (k = 48), as shown in Table 4. As discussed below, one limitation of this analysis is that not all studies that reported significant length effects reported the necessary values to contribute effect sizes to the length analysis. We followed the procedures for calculating effect sizes and homogeneity statistics as described previously.

Because of the interdependence between the effect sizes at the nonword lengths (i.e., many of the same samples contributed effect sizes at several lengths), we do not interpret the overall effect size, which would be biased

Table 4. Exploratory meta-analysis for one-, two-, three-, and four-syllable items.

	One-syllable items		Two-syllable items		Three-syllable items		Four-syllable items	
Article	*d*	CI	*d*	CI	*d*	CI	*d*	CI
Archibald & Gathercole (2006) CNRep			0.81	−0.05–1.61	1.22	0.31–2.05	1.34	0.41–2.17
Coady et al. (2006)					1.05	0.37–1.69	1.44	0.72–2.11
Dollaghan & Campbell (1998)	0.67	0.02–1.29	1.06	0.38–1.70	1.52	0.79–2.19	1.56	0.82–2.23
Ellis Weismer et al. (2000)	0.42	0.18–0.66	0.40	0.15–0.64	0.51	0.27–0.76	0.57	0.33–0.82
Evans et al. (2005)								
Older	0.34	−0.38 to 1.04	0.92	0.15–1.64	1.12	0.33–1.85	1.62	0.77–2.38
Evans et al. (2005)								
Younger	−0.05	−0.51–0.41	0.45	−0.01–0.91	0.73	0.25–1.19	1.12	0.62–1.60
Gathercole & Baddeley (1990)	1.58	0.04–2.82						
Gray (2003)			1.66	0.95–2.32	2.69	1.84–3.45	2.33	1.53–3.05
Horohov & Oetting (2004)	1.27	0.53–1.95	1.13	0.40–1.80	1.63	0.85–2.35	1.36	0.61–2.05
Kamhi et al. (1988)	2.95	1.59–4.07						
Marton & Schwartz (2003)			0.43	−0.36–1.19	1.32	0.43–2.12	1.37	0.48–2.18
Montgomery (1995b)	0.72	−0.08–1.47	0.43	−0.34–1.18	1.82	0.88–2.66	2.37	1.33–3.27
Montgomery (2004)	0.80	−0.05–1.61	0.40	−0.42–1.20	1.78	0.78–2.65	1.95	0.93–2.85
Munson et al. (2005)								
Low probability only					1.13	0.35–1.84	1.12	0.35–1.83
Rodekohr & Haynes (2001)								
African American	0.59	−0.33–1.46	0.17	−0.72–1.04	1.15	0.16–2.04	0.66	−0.26–1.53
Rodekohr & Haynes (2001)								
White	1.06	0.08–1.95	0.00	−0.88–0.88	1.42	0.39–2.34	1.29	0.28–2.19

Note. Effect size (*d*) is the unweighted, uncorrected effect size.

because of the nonindependent samples. We can still examine the effect sizes for each nonword length, as these represent independent samples. The total nonhomogeneity was significant (H_T = 159.23, p < .05), thereby justifying the exploratory moderator analyses to examine effect sizes as a function of nonword length. Effect size magnitude increased with nonword length, such that one- and two-syllable words showed medium effect sizes (d = .52 and d = .57, respectively), and three- and four-syllable words showed large effect sizes (d = .95 and d = 1.05). The test of between-groups homogeneity was significant (H_B = 29.47, p < .05), indicating that nonword length was significantly related to the magnitude of effect sizes. The within-group homogeneity test was also significant (H_w = 129.76, p < .05), indicating that there was significant nonhomogeneity remaining among the effect sizes.

To examine model fit, we ran a multiple regression with nonword length predicting effect size magnitude, using the inverse of effect size variance as a case weight. In one step, we contrasted one- and two-syllable nonwords (each coded −1) to three- and four-syllable nonwords (each coded 1). The contrast was significant (β = .42, z = 10.06, p < .05, CI = .34–.51), indicating that the effect sizes for the three- and four-syllable nonwords were significantly greater than the effect sizes for the one- and two-syllable nonwords. The model explained 18.0% of the variance in effect sizes. Model fit was rejected (H_E = 130.57, p < .05), indicating that factors in addition to nonword length were responsible for the nonhomogeneity of the sample of effect sizes. Because of the small number of studies in the analysis, we did not pursue additional moderators.

Summary of results. In the overall analysis, children with and without SLI exhibited a large and statistically significant difference in performance. There was significant variability in the distribution of effect sizes. A moderator analysis revealed that the type of nonword repetition measure used explained a significant proportion of the variability, yet there was remaining unexplained variability. The CNRep yielded a significantly larger effect size than the three- to four-syllable nonword set and the NRT and a similar magnitude effect size to the Montgomery nonword set (see Table 3). The second moderator, age of the SLI sample, was not significantly related to variability in effect sizes. The final analysis explored a subset of the studies contributing effect sizes broken down by nonword length. Children with SLI were found to perform significantly below children with NL at all nonword lengths, with greater deficits on three- to four-syllable items than on one- to two-syllable items.

Discussion

This meta-analysis examined the magnitude of the difference in nonword repetition performance between children with SLI and children with NL. Our analysis revealed a very large overall weighted mean effect size, d = 1.27, which indicates that children with SLI scored an average of 1.27 standard deviations below typically developing children (from 22 CAM or nonverbal IQ-matched and 2 LM samples). Nonword repetition mimics aspects of word learning, a process that many children with language impairments find challenging (e.g., Oetting, Rice, & Swank, 1995). Children hearing a new phonological sequence must perceive and encode the sequence appropriately. They must also hold the sequence in a temporary memory store with a robust enough representation to support further processing of the sound sequence, articulation, and connection to meaning. Children with SLI may be impaired at any point, or at multiple points in this process. Such impairments could contribute to difficulty learning words as well as grammatical systems (reviewed in Montgomery, 2002).

Although children with SLI showed a substantial deficit in nonword repetition, the magnitude of this deficit varied significantly across studies. Of the effect sizes in the analysis, children with SLI performed between 0.62 (Washington & Craig, 2004) and 4.34 standard deviations (Gathercole & Baddeley, 1990) below children with NL. The moderator analyses helped to explain this variability and addressed important unanswered questions concerning the nonword repetition performance of children with SLI.

The first moderator analysis addressed the following question: Are all versions of the nonword repetition task interchangeable? The results indicate that measures of nonword repetition are not interchangeable. The magnitude of the nonword repetition deficit in children with SLI was significantly associated with the type of task used to assess it. The CNRep yielded the largest effect size, followed by the Montgomery nonword set, the three- to four- syllable sets, and the NRT. While other unreported variables also contributed to the variability, finding that the measures yielded significantly different effect sizes indicates that the characteristics of nonword repetition tasks require deeper investigation. Understanding the nature of the nonword repetition task is important because of the growing application of the task in research focused on shaping both theories of language impairments and the practice of identifying children with language impairments. The CNRep and the NRT yielded the most extreme effect sizes, they are often grouped together in literature reviews, and they have been adopted by several research groups, including those involved in large-scale genetic studies (e.g., the CNRep used by the SLI Consortium, 2002). Therefore, the comparison of these measures is of particular interest.

The CNRep yielded a significantly higher estimate of the size of the nonword repetition deficit than the NRT. Similarly, Archibald and Gathercole (2006) concluded that children with SLI exhibit greater deficits on the CNRep than on the NRT, based on comparisons of children with SLI to CAM and LM groups. (However, Table 2 shows that when performance is not corrected for nonverbal cognition, the effect size of the NRT is somewhat larger.) In addition, Gray (2003, 2004) used both the CNRep and the NRT with children with similar characteristics (two samples of children of similar ages and similar identification procedures). Gray found a larger effect size for the sample tested with the CNRep (see Table 2). We propose that differences in design characteristics, such as those listed in Table 1, contributed to the discrepant effect sizes seen across tasks (see also Archibald & Gathercole, 2006; Coady et al., 2006).[3]

Several characteristics of the CNRep appear to favor the very large effect size of the measure. First, the high proportion of long nonwords may tax the phonological working memory of children with SLI to a greater extent than children with NL, and to a greater extent than measures with fewer long words (i.e., the NRT). This proposal is supported by the findings of several studies as well as our nonword length analysis. Similarly, the already vulnerable language processing skills of children with SLI may be more affected by detrimental fatigue and interference effects produced by long lists of nonwords than children with NL (see also the large effects for the Montgomery nonword sets).

Articulatory complexity is another factor that may contribute to differences in performance across nonword repetition tasks. Children with SLI may fall more behind children with NL on tasks that provide greater opportunity for articulation errors (Bishop et al., 1996; Stark & Blackwell, 1997; but see Edwards & Lahey, 1998). The CNRep (as well as the Montgomery nonword set) provides such opportunities by including consonant clusters, whereas the NRT excludes clusters and late developing consonants. Thus, the complexity of the CNRep appears to promote higher error rates for children with SLI and therefore larger effect sizes than the NRT. Performance on the CNRep may also draw a larger contribution from motor planning or articulation skills than measures with minimal articulatory difficulty.

Another possible source of differences across measures is the scoring method used. The CNRep and several other measures (see Table 1) typically use an all-or-nothing scoring method in which repetitions are scored as the number of items fully correct, a method that facilitates online scoring. The NRT is typically scored as the percentage of phonemes produced correctly. Previously we suggested that if children with SLI are more likely to make small errors in repetitions, a percentage-of-phonemes-correct scoring method may produce smaller effect sizes because the children do not have to produce items without error to receive credit.

To explore the effects of scoring method, we examined the nonword repetition performance of a group of children with SLI and a CAM group (n = 68) who were recently tested with the NRT in our laboratory. Their responses were scored using both methods. The total-number-correct scoring method produced lower scores for both children with SLI and children with NL compared with the percentage-of-phonemes-correct scoring (see Table 5 for values). In fact, the effect size for the total-number-correct scoring was lower (d = .48) than that of the percentage-of-phonemes-correct scoring (d = 1.17), due in part to the larger pooled standard deviation in the total-number-correct scoring. This preliminary analysis suggests that both groups of children receive lower nonword repetition scores by using the total-number-correct scoring method but that it should not contribute to larger effect sizes.

The wordlikeness of the test items is another factor that could affect the magnitude of the nonword repetition deficit across tasks. The CNRep (as well as some items in the Montgomery set) is described as consisting of items with mixed wordlikeness, based on measures of phonotactic probability and wordlikeness ratings (ratings of wordlikeness correlate with measures of phonotactic probability, as in Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997). However, even some of the low-wordlike nonwords on the CNRep contain real English words and affixes. In contrast, the items on the NRT are minimally wordlike.

It is not clear how wordlikeness should relate to the magnitude of the nonword repetition deficit in children with SLI. There is mixed evidence concerning wordlikeness effects in this group. To apply Archibald and Gathercole's (2006) proposal, if children with SLI are not readily able to recruit lexical support when repeating novel words, the high wordlikeness of the CNRep may contribute to its very large effect size. In contrast, Munson

Table 5. Comparison of scoring methods for performance on the NRT.

Scoring method	SLI M (SD)	CAM M (SD)
Number correct scoring (presented as percentage correct out of 16)	49.2 (16.2)	57.2 (17.2)
Percentage of phonemes correct	75.2 (10.8)	86.0 (7.5)
Age (in years)	9.6 (1.5)	9.6 (1.8)

[3] The CNRep has been primarily used with British children (e.g., Bishop et al., 1996; Briscoe et al., 2001; Conti-Ramsden & Hesketh, 2003), and the NRT has been used with American children (e.g., Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000). However, because most nonword repetition research has not been presented as specifically applying to children of a particular nationality, we find little reason to believe the nonword repetition differences across measures are due to differences in nationality or English dialect.

et al. (2005) found that compared with CAM children, children with SLI were at a greater disadvantage for low phonotactic probability words than for high probability words. This phonotactic probability effect suggests that when lexical support is not readily available (as with nonwords that are dissimilar to real words), the phonological memory skills of children with SLI provide an insufficient foundation for the task. However, there is also evidence that children with SLI show a similar pattern of facilitation for high-wordlike nonwords and difficulty with low-wordlike nonwords compared with CAM children (i.e., there is no interaction of wordlikeness and language status; Briscoe et al., 2001; Coady et al., 2006).

While the conclusions of previous studies of wordlikeness do not yet present a clear rationale for why nonword repetition measures produce different magnitude effect sizes, the wordlikeness effects do present a clear case for the ways that the characteristics of nonword repetition measures matter. The low wordlikeness of the NRT likely taps phonological skills that are less dependent (but probably not independent) on stored word knowledge, whereas the CNRep likely recruits support from both phonological and articulation skills as well as the lexical repertoire.

In summary, the different effect sizes associated with the various nonword repetition measures indicate that they are not interchangeable. The differences in effect sizes may reflect test characteristics that differentially tax children with SLI, such as nonword length, articulatory complexity, and wordlikeness. Future investigations of the effects of manipulating such characteristics have potential to reveal the nature of the impairments that contribute to SLI.

A second moderator analysis examined the following question: Are there age differences in the magnitude of the nonword repetition deficit in children with SLI? The moderator analysis reported no significant relation between effect size and the age of children in the SLI samples. Age did not help explain the variability in effect sizes. This finding was a null finding; therefore, any interpretation must be taken with caution. The conclusions that can be drawn are limited because of the cross-sectional nature of the comparison and because the analysis collapsed across studies using different designs and different measures. Our age analysis highlights the need for future studies that are specifically designed to address the developmental trajectory of nonword repetition performance in SLI, as we know very little about the development of processing-related impairments.

The pattern of results is consistent with the hypothesis that the magnitude of the nonword repetition deficit in children with SLI remains relatively stable across ages. Children with NL and with SLI may maintain similar differences in performance as the performance of both groups improves. The result is also consistent with the lasting nonword repetition deficits seen in children with resolved language impairments (Bishop et al., 1996). There is no support for a clear increase or decrease in the advantage for children with NL that could be attributed to differential changes in reliance on phonological working memory across the groups or to differential gains in access to lexical knowledge. The consistency of the deficit indicates that measures of nonword repetition may be useful for identifying children with SLI across development, from early school age to adolescence.

We used an exploratory meta-analysis to address the third question: Are there circumstances under which the nonword repetition skills of children with SLI are comparable to those of children with NL? Several studies have reported that children with SLI do not exhibit statistically significant impairments for repeating short nonwords (e.g., Dollaghan & Campbell, 1998) but exhibit consistent and large impairments for long nonwords. However, the lack of significant differences between children with and without SLI on short nonwords may be due in part to a lack of statistical power. Studies using large samples (Ellis Weismer et al., 2000) or long lists of one-syllable items (Kamhi et al., 1988, Gathercole & Baddeley, 1990) have found that children with SLI have substantial difficulty repeating short nonwords. In the exploratory meta-analysis, we analyzed performance at each nonword length by combining data from all studies that reported sufficient data to calculate effect sizes. A limitation of the analysis was that length effect size information was unavailable for most studies using the CNRep, although some studies reported that children with SLI performed increasingly worse on longer nonwords on this measure (e.g., Bishop et al., 1996). As discussed previously, the CNRep produced many of the largest effect sizes in the analysis and includes the longest nonwords. Therefore, our length analysis may underestimate the overall magnitude of the length effects. However, the analysis showed that children with SLI were impaired at all nonword lengths, with significantly greater deficits for longer nonwords (three and four syllables) than for shorter nonwords (one and two syllables).

While the increased difficulty of the longer nonwords was not surprising, a key finding emerged. Children with SLI had significant difficulty repeating even one-syllable sequences. While this deficit may not be clinically significant (the performance of children with SLI is likely to be more different from their NL peers on longer nonwords), it possesses theoretical relevance. Limitations in the phonological working memory capacity of children with SLI have been argued to be a possible source of their pervasive language impairments (Gathercole & Baddeley, 1990; Montgomery, 2002, 2004). Montgomery (2002) and Gathercole and Baddeley have indicated that the increase in the difference between children with and without SLI and the increase in difficulty from shorter to longer nonwords reflect the limitations in phonological working memory resources. How does a deficit in repeating even very simple nonwords fit within a limited phonological memory capacity account?

A deficit repeating one- and two-syllable items may implicate mechanisms other than, or in addition to, limited phonological working memory capacity. Perhaps, at least, some children with SLI have difficulty with the initial phonological perception or encoding of word forms, or have degraded lexical representations that contribute to the difficulty of the task. For example, while a one-syllable novel word may not be sufficient to overwhelm working memory capacity, children who fail to accurately encode or represent the sound sequence may have trouble repeating it. Longer nonwords may produce compounding effects for children with SLI if they lack support from the initial phonological representations or from associations with lexical knowledge, and the nonwords overwhelm phonological working memory capacity as well.

To return to the question of whether children with SLI ever appear unimpaired on nonword repetition tasks, the current evidence for nonword length suggests that they are impaired at all item lengths. There may be conditions, such as with short and high-wordlikeness items, under which children with SLI exhibit unimpaired performance compared with children the same age. Regardless, the fragility of their performance on the task indicates that there exists a substantial deficit in those skills tapped by repeating novel words.

The present meta-analyses and moderator analyses have provided valuable information regarding patterns in the nonword repetition performance of children with SLI and have pointed to topics that merit additional attention. However, the significant nonhomogeneity remaining after the analyses indicates that the types of nonword repetition measures used across experiments and nonword length are not the only factors affecting the magnitude of effect sizes. While the age of the SLI sample did not relate to differences in effect sizes, there are other sample-based characteristics that could affect the magnitude of the difference between children with and without SLI.

The different means of identifying children with SLI used across studies present an additional factor that could affect effect size magnitude. Studies used a variety of measures and language skill diagnostic criteria, including intervention status (e.g., Dollaghan & Campbell, 1998) and performance under 1.3 standard deviations below the mean on two of three receptive and expressive language subtests (Montgomery, 2004). We investigated the possibility of testing diagnostic criteria or the severity of the SLI groups as moderators. However, we were somewhat surprised to see that when data were available to examine children with SLI on comparable language measures, the samples looked highly similar across studies. There was not sufficient variability or, for many studies, sufficient information available to examine whether children with SLI of differing severity or subtype (e.g., expressive–receptive vs. expressive impairments only) also differ in the magnitude of their nonword repetition deficit. This is an important topic for future research; different means of identifying children with SLI may tap children with somewhat different underlying impairments, including perhaps a range of nonword repetition deficits.

Another important difference across studies is the type of control group used. In our analyses, we examined the performance of children with SLI compared with a CAM group when such a comparison was possible (23 of 25 samples). We selected the CAM group to promote consistency in the comparisons because studies using LM control groups used a variety of measures of language skills as the basis for matching (e.g., receptive and expressive vocabulary, receptive syntax). It was not feasible to analyze the type of control group as a moderator because including data from both control groups would violate the meta-analysis assumption of independent effect sizes. In addition, the distribution of CAM and LM control groups is uneven across nonword repetition measures. Four studies using the CNRep and one study reporting NRT scores used both CAM and LM samples. To break down the effect sizes of the CNRep to include all possible LM samples (and exclude the CAM comparison for the same study) would weaken the integrity of the nonword repetition measure comparison, a comparison that holds theoretical and practical significance.

However, it is important to consider how the effect sizes differed when studies use CAM versus LM groups. Van der Lely and Howard (1993) indicated that comparisons against CAM groups can be less informative than LM comparisons because children with SLI are known to fall behind children the same age on most language-related tasks. However, Plante, Swisher, Kiernan, and Restrepo (1993) pointed out that there are significant limitations in the use of LM samples as well. As seen in Table 2, in seven of the eight studies using both CAM and LM control groups, effect sizes were smaller with the LM group. Yet, for five of the eight LM samples, the effect sizes remained large or very large (above .9). It is not currently possible to clarify the effects of the type of control group, or the effects of different types of language matching. The small and variable amount of data available indicate that it is possible that, overall, children with SLI show poorer nonword repetition performance in comparison to younger, LM children in addition to children their own age.

This meta-analysis demonstrates that characteristics of nonword repetition tasks matter for the magnitude of the SLI deficit found. It also suggests future directions for understanding how nonword repetition measures recruit varying contributions from phonological working memory, lexical knowledge, and articulatory motor planning. The results may also affect the application of nonword repetition as an identifier of children with SLI. The analysis of age patterns indicates that nonword repetition holds promise for identifying children with SLI at many ages. The differences across measures indicate that although all measures find statistically significant differences between children with and without SLI, the measures are not interchangeable. It is possible that they are not all tapping the same sort of deficit in children with SLI and may even identify somewhat different (though overlapping) groups of children with language impairments.

The effect size analysis and the discussion of task characteristics indicate that the CNRep may differentiate some groups of children with and without SLI to a greater degree than the NRT. The CNRep could tap subtle articulation difficulties and a lack of support from lexical or sublexical mediation, in addition to problems with phonological memory resources. If this is the case, it suggests that measures that compound influences on performance promote the largest differences in performance between children with and without language impairments. In contrast, the NRT, because of its low wordlikeness and complexity, may be more closely and specifically linked to basic phonological processing or memory skills than the CNRep. Clearly, the skills that nonword repetition recruits are similar across measures; all require children to repeat novel sound sequences. However, this meta-analysis indicates that as researchers and clinicians apply nonword repetition to measure underlying skills in SLI and to identify children with language impairments, it is important to consider the nature of the task that is used.

Acknowledgments

This research was supported in part by Research Grant R01 DC005650 from the National Institute on Deafness and Other Communication Disorders (NIDCD) to Julia Evans and by Grant F31 DC07277 from NIDCD to Katharine Graf Estes. We would like to thank Jeffry Coady and Janet Hyde for their helpful comments and Lisbeth Simon, Kristin Ryan, and Rebecca Porwoll for their assistance gathering data. We would also like to thank James Montgomery, Benjamin Munson, Shelley Gray, and Jeffry Coady for graciously providing additional unpublished findings to include in the analysis. A previous version of these findings was presented as a poster at the June 2004 Symposium on Research in Child Language Disorders, Madison, WI.

References

Archibald, L. M. D., & Gathercole, S. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research, 49,* 970–983.

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36,* 189–208.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158–173.

Bishop, D. V. M. (1992). The underlying nature of specific language impairment. *Journal of Child Psychology and Psychiatry, 33,* 1–64.

Bishop, D. V. M. (2000). The Mac Keith Lecture 1999: How does the brain learn language? Insights from the study of children with and without language impairment. *Developmental Medicine and Child Neurology, 42,* 133–142.

Bishop, D. V. M., Bishop, S. J., Bright, P., James, C., Delaney, T., & Tallal, P. (1999). Different origin of auditory and phonological processing problems in children with language impairment: Evidence from a twin study. *Journal of Speech, Language, and Hearing Research, 42,* 155–168.

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 37,* 391–403.

Botting, N., & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International Journal of Language and Communication Disorders, 36,* 421–432.

Bowey, J. A. (1996). On the association between phonological memory and receptive vocabulary in five-year-olds. *Journal of Experimental Child Psychology, 63,* 44–78.

Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics, 22,* 441–469.

Briscoe, J., Bishop, D. V. M., & Norbury, C. F. (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42,* 329–340.

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment: Processing-dependent measures. *Journal of Speech, Language, and Hearing Research, 40,* 519–525.

Coady, J. A., & Evans, J. L. (in press). The uses and interpretations of nonword repetition tasks in children with and without specific language impairments. *International Journal of Language and Communication Disorders*.

Coady, J. A., Evans, J. L., & Kluender, K. R. (2006). *The role of phonotactic frequency in nonword repetition by children with specific language impairments*. Manuscript in preparation.

Coady, J. A., Evans, J., Kluender, K. R., Mainela-Arnold, E., & Ryan, K. (2003, June). *Nonword repetition in children with E–SLI and ER–SLI*. Poster presented at the 24th Annual Symposium on Research in Child Language Disorders, Madison, WI.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conti-Ramsden, G. (2003). Processing and linguistic markers in young children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research, 46,* 1029–1037.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42,* 741–748.

Conti-Ramsden, G., & Hesketh, A. (2003). Risk markers for SLI: A study of young language-learning children. *International*

Journal of Language and Communication Disorders, 38, 251–263.

Dollaghan, C. (1998). Spoken word recognition in children with and without specific language impairment. *Applied Psycholinguistics, 19,* 193–207.

Dollaghan, C. A., Biber, M. E., & Campbell, T. F. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics, 16,* 211–222.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1136–1146.

Durlak, J. A. (2003). Basic principles of meta-analysis. In M. C. Roberts & S. S. Ilardi (Eds.), *Handbook of research methods in clinical psychology* (pp. 196–209). Malden, MA: Blackwell.

Edwards, J., & Lahey, M. (1998). Nonword repetitions of children with specific language impairment: Exploration of some explanations for their inaccuracies. *Applied Psycholinguistics, 19,* 279–309.

Ellis Weismer, S., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research, 43,* 865–878.

Evans, J. L., Graf Estes, K., Coady, J. A., Ryan, K., & Simon, L. (2005). [Lab report on nonword repetition performance in children with SLI]. Unpublished data.

Farmer, M. (2000). Language and social cognition in children with specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41,* 627–636.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23,* 83–94.

Gathercole, S. E., & Adams, A. M. (1993). Phonological working memory in very young children. *Developmental Psychology, 29,* 770–778.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children. *Journal of Memory and Language, 28,* 200–213.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29,* 336–360.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A. M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology, 13,* 65–77.

Gathercole, S. E., Tiffany, C., Briscoe, J., Thorn, A., & ALSPAC Team. (2005). Developmental consequences of poor phonological short-term memory function in childhood: A longitudinal study. *Journal of Child Psychology and Psychiatry, 46,* 598–611.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28,* 887–898.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory, 2,* 103–127.

Goffman, L. (2004). Kinematic differentiation of prosodic categories in normal and disordered language development. *Journal of Speech, Language, and Hearing Research, 47,* 1088–1102.

Gray, S. (2003). Diagnostic accuracy and test–retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders, 36,* 129–151.

Gray, S. (2004). Word learning by preschoolers with specific language impairment: Predictors and poor learners. *Journal of Speech, Language, and Hearing Research, 47,* 1117–1132.

Hedges, L. V., & Becker, B. J. (1986). Statistical methods in the meta-analysis of research and gender differences. In J. S. Hyde & M. C. Lynn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 14–50). Baltimore: Johns Hopkins University Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504.

Horohov, J. E., & Oetting, J. (2004). Effects of input manipulations on the word learning abilities of children with and without specific language impairment. *Applied Psycholinguistics, 25,* 43–65.

Kamhi, A. G., & Catts, H. W. (1986). Toward an understanding of developmental language and reading disorders. *Journal of Speech and Hearing Research, 51,* 337–347.

Kamhi, A. G., Catts, H. W., Mauer, D., Apel, K., & Gentry, B. F. (1988). Phonological and spatial processing abilities in language- and reading-impaired children. *Journal of Speech and Hearing Disorders, 53,* 316–327.

Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Marton, K., & Schwartz, R. G. (2003). Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 46,* 1138–1153.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91,* 3–19.

Montgomery, J. W. (1995a). Examination of phonological working memory in specifically language-impaired children. *Applied Psycholinguistics, 16,* 355–378.

Montgomery, J. W. (1995b). Sentence comprehension in children with specific language impairment: The role of phonological working memory. *Journal of Speech and Hearing Research, 38,* 187–199.

Montgomery, J. W. (2002). Understanding the language difficulties of children with specific language impairments: Does verbal working memory matter? *American Journal of Speech-Language Pathology, 11,* 77–91.

Montgomery, J. W. (2003). Working memory and comprehension in children with specific language impairment: What we know so far. *Journal of Communication Disorders, 36,* 221–231.

Montgomery, J. W. (2004). Sentence comprehension in children with specific language impairment: Effects of input rate and phonological working memory. *International Journal of Language and Communication Disorders, 39,* 115–133.

Montgomery, J. W., & Windsor, J. (2004). *Examining the language performance of children with SLI: Contributions of phonological working memory and speed of processing*. Manuscript submitted for publication.

Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without language impairments. *Journal of Speech, Language, and Hearing Research, 48,* 1033–1047.

Oetting, J. B., Rice, M. L., & Swank, L. (1995). Quick incidental learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech, Language, and Hearing Research, 38,* 434–445.

Plante, E., Swisher, L., Kiernan, B., & Restrepo, M. A. (1993). Language matches: Illuminating or confounding? *Journal of Speech and Hearing Research, 36,* 772–776.

Robey, R. R., & Dalebout, S. D. (1998). A tutorial on conducting meta-analyses of clinical outcome research. *Journal of Speech and Hearing Research, 41,* 1227–1241.

Rodekohr, R. K., & Haynes, W. O. (2001). Differentiating dialect from disorder: A comparison of two processing tasks and a standardized language test. *Journal of Communication Disorders, 34,* 255–272.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86,* 638–641.

Shriberg, L. D., & Kwiatkowski, J. (1994). Developmental phonological disorders: I. A clinical profile. *Journal of Speech and Hearing Research, 37,* 1100–1126.

SLI Consortium. (2002). A genomewide scan identifies two novel loci involved in specific language impairment. *American Journal of Human Genetics, 70,* 384–398.

SLI Consortium. (2004). Highly significant linkage to SLI1 locus in an expanded sample of individuals affected by specific language impairment. *American Journal of Human Genetics, 74,* 1225–1238.

Snowling, M. J., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics, 12,* 369–373.

Stark, R. E., & Blackwell, P. B. (1997). Oral volitional movements in children with language impairments. *Child Neuropsychology, 3,* 81–97.

Thomas, M., & Karmiloff-Smith, A. (2005). Can developmental disorders reveal component parts of the human language faculty? *Language Learning and Development, 1,* 65–92.

van der Lely, H. K. J., & Howard, D. (1993). Children with specific language impairment: Linguistic impairment or short-term memory deficit? *Journal of Speech and Hearing Research, 36,* 1193–1207.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for processing of spoken nonsense words. *Language and Speech, 40,* 47–62.

Washington, J. A., & Craig, H. K. (2004). A language screening protocol for use with young African American children in urban settings. *American Journal of Speech-Language Pathology, 13,* 329–340.

Zamuner, T. S., Gerken, L., & Hammond, M. (2004). Phonotactic probabilities in young children's speech production. *Journal of Child Language, 31,* 515–536.

Received June 30, 2005

Revision received December 7, 2005

Accepted May 24, 2006

DOI: 10.1044/1092-4388(2007/015)

Contact author: Katharine Graf Estes, Department of Psychology, Waisman Center, University of Wisconsin—Madison, 1500 Highland Avenue, Madison, WI 53705-2280. E-mail: kmgraf@wisc.edu.

Nicole M. Else-Quest is now at Villanova University, Villanova, PA.