

Nonword Repetition and Child Language Impairment

Chris Dollaghan

University of Pittsburgh
Pittsburgh, PA

Thomas F. Campbell

University of Pittsburgh
Children's Hospital of Pittsburgh

A brief, processing-dependent, nonword repetition task, designed to minimize biases associated with traditional language tests, was investigated. In Study 1, no overlap in nonword repetition performance was found between a group of 20 school-age children enrolled in language intervention (LI) and a group of 20 age-matched peers developing language normally (LN). In Study 2, a comparison of likelihood ratios for the nonword repetition task and for a traditional language test revealed that nonword repetition distinguished between children independently identified as LI and LN with a high degree of accuracy, by contrast with the traditional language test. Nonword repetition may have considerable clinical utility as a screening measure for language impairment in children. Information on the likelihood ratios associated with all diagnostic tests of language is badly needed.

KEY WORDS: child language impairment, language assessment, nonword repetition, test bias, language screening

The literature on child language impairment is replete with descriptions of tasks on which groups of children with impaired language (LI) perform less well than their peers developing language normally (LN). Such group differences have most often been used in support of particular hypotheses concerning the underlying causal mechanisms of child language impairment. There has been surprisingly little attention to the question of the diagnostic utility of the many differences observed between children with LI and LN, or the extent to which group differences in performance can be used to accurately classify individual children with LI or LN.

This question is particularly germane because the criteria used to define a child with LI in the first place usually involve performance below some cut-off score on one or more norm-referenced tests (e.g., Tomblin, Records, & Zhang, 1996), possibly in addition to a subjective judgment from one of the child's family members or teachers (Paul, 1995). We recently argued that such norm-referenced tests are inherently biased against test-takers from minority backgrounds because these tests depend so heavily on experiential history generally, and on vocabulary knowledge specifically (Campbell, Dollaghan, Needleman, & Janosky, 1997). We reported evidence that processing-dependent measures, designed to be equal in familiarity to all test-takers regardless of their language knowledge, are less biased against children from minority backgrounds than are such knowledge-dependent measures. We suggested that processing-dependent measures offer a better way to distinguish between children whose poor performance reflects fundamental language processing deficits and children whose poor performance can be attributed to their differing experiential histories.

In the present paper, we consider the diagnostic utility of one such processing-dependent measure, nonsense word repetition. There is considerable evidence that groups of children with language impairment repeat nonsense words less accurately than do their peers developing language normally (Gathercole & Baddeley, 1990a, 1993; Montgomery, 1995); in fact, Bishop, North and Donlan (1996) recently suggested that nonword repetition provides a phenotypic marker for some forms of developmental language impairment. However, the bulk of the existing evidence of a nonword repetition deficit in children with LI has come from nonword stimuli that have not been designed to minimize the potential influence of prior language knowledge on repetition performance. Unless nonwords are designed to ensure that they are equally unfamiliar to children with LI and LN, the poor repetition of children with LI could be attributed to their reduced language knowledge rather than to a fundamental psycholinguistic deficit (Dollaghan, Biber, & Campbell, 1993, 1995; Gathercole, Willis, Emslie, & Baddeley, 1991; Snowling, Chiat, & Hulme, 1991). This requires that nonword repetition tasks be designed such that neither the nonwords nor their constituent syllables correspond to lexical items; further, the predictability of individual phonemes within the nonwords should be minimized. In addition, nonwords ideally would include phonemes that are acquired early in development (so that poor repetition performance cannot be attributed to articulation deficits) and are acoustically salient (for the sake of both the child and the transcriber). Finally, the presentation of the nonwords should be standardized to ensure that stimuli are presented with consistent rate, accuracy, and intonation.

In Study 1, we designed a set of nonword stimuli meeting these criteria, and compared the accuracy with which groups of age-matched school-age children with and without LI repeated them. In Study 2, we examined the diagnostic accuracy of this nonword task in a larger ($N = 85$) sample of children, by comparing the likelihood ratios for several levels of nonword repetition performance with the likelihood ratios for several levels of performance on a norm-referenced test of language abilities.

Likelihood ratios represent a means of evaluating the clinical value of one or more diagnostic measures (Sackett, Haynes, Guyatt, & Tugwell, 1991). Like the more familiar metrics of sensitivity and specificity, likelihood ratios express a measure's success in distinguishing among those members of a population who are affected by a condition (in this case, language impairment), and those without a disorder. Unlike sensitivity and specificity, however, likelihood ratios are affected minimally by variations in the prevalence rate for a disorder. In addition, they can be calculated for several levels of performance on a measure to provide additional

evidence concerning the diagnostic significance of a given client's performance. By applying the likelihood ratio strategy to a processing-dependent and a traditional knowledge-dependent measure of language impairment, our intent is to illustrate the likelihood-ratio approach generally, while examining the diagnostic efficiency of the present nonword repetition task.

Study 1

Purpose

The purpose of Study 1 was to examine the extent to which a nonword repetition task constructed to minimize the influence of subjects' previous language knowledge differentiated age-matched children with and without language impairment.

Method

Participants

Participants in Study 1 were 40 children between the ages of 6;0 and 9;9 (years;months), drawn from a larger, ongoing study of child language impairment. Twenty of the children had been diagnosed by an ASHA-certified school speech-language pathologist as having a language impairment and were enrolled in language intervention in an urban public school setting. Each of the other 20 children, developing language normally and not enrolled in language or speech intervention by report of the school speech-language pathologist, was matched on age (± 3 months) to one of the children with LI. The mean age of both groups was 7;10. There were 14 males and 6 females in each group; all were native speakers of English. Both groups were similar in parent-reported ethnic diversity, with the number of African American, White, and other participants in the LN and LI groups being 13 and 12, 4 and 5, and 3 and 3, respectively.

We used intervention status rather than one or more norm-referenced test scores as the *gold standard* for the condition of LI for two reasons. First, the issue of how to define language impairment continues to be a contentious one; as noted by a number of investigators (e.g., Aram, Morris, & Hall, 1993; Dunn, Flax, Sliwinski, & Aram, 1996; Lahey, 1990; Tomblin, Records, & Zhang, 1996) there is no consensus on a gold standard for the diagnosis of language impairment. Second, as discussed earlier, evidence suggests that existing norm-referenced tests are biased against children whose backgrounds differ from those of the majority of the test's normative sample (Campbell et al.,

1997; Fazio, Naremore, & Connell, 1996; Fischer et al., 1996). Intervention status ostensibly reflects the degree of concern that speech-language pathologists, teachers, and parents have about an individual child's language skills, within a particular school's social and economic context. Such concerns can originate from a variety of sources, singly or in combination, including teacher observations and referral, screening measures, norm-referenced tests, and criterion-referenced measures such as language sample analysis. Factors other than a child's overtly measured language performance may enter into the decision to enroll him or her in language intervention, such as the clinician's experience and theoretical perspective and the policies of the institution within which he or she works (Records & Tomblin, 1994). These multifaceted clinical judgments have more credibility as a gold standard for the condition of language impairment than do any existing norm-referenced tests for children from minority backgrounds.

Although intervention status was used as the sole criterion for language impairment, it is of interest to compare the performance of the resulting groups on some commonly used norm-referenced tests (Table 1) that had been administered to each participant in the course of the larger study. Not surprisingly, group comparisons (with alpha level set at $p < .05$) revealed that the standard score equivalents, or z -scores, of the children with LI were significantly lower than those of the children with LN on both the Peabody Picture Vocabulary Test–Revised (Dunn & Dunn, 1981) and the Spoken Language Quotient of the Test of Language Development–2 (Hammill & Newcomer, 1988; Newcomer & Hammill, 1988). There was also a significant group difference on the Test of Nonverbal Intelligence–Revised (TONI-R) (Brown, Sherbenou, & Johnsen, 1990). This finding of a group mean difference in nonverbal reasoning in favor of the children developing language normally is consistent with evidence (Leonard, 1987) that even when

groups of children with language impairments are selected to perform within normal limits on nonverbal measures of cognition, their group mean score is likely to be significantly lower than that of age-matched children developing language normally.

Stimuli

Sixteen nonwords (Table 2), four at each of four syllable lengths (one, two, three, and four syllables), were constructed for the present study. All nonwords began and ended with consonants (Cs); they contained no consonant clusters. Thus, one-syllable nonwords were CVCs; two-syllable nonwords CVCVCs; three-syllable nonwords CVCVCVCs; and four-syllable nonwords CVCVCVCVCs, for a total of 96 phonemes over the entire nonword set. In addition, nonwords met the following specific constraints:

1. To ensure that nonword repetition would not be affected by a subject's vocabulary knowledge, nonwords were constructed such that none of their individual syllables (CV or CVC) corresponded to an English word.

2. To minimize the articulatory difficulty of the repetition task, enabling the inference that errors resulted from a lack of recall of target phonemes, rather than from an inability to produce them, the nonwords were constructed to exclude the consonants described by Shriberg and Kwiatkowski (1994) as the "Late Eight" (i.e., /s, z, l, r, ʃ, ʒ, θ, ð/), as well as consonant clusters.

3. The nonwords were constructed to contain only tense vowels, for two reasons. First, being longer in duration than lax vowels, tense vowels are inherently less susceptible to being reduced to schwa. Thus, by contrast with lax vowels, errors on tense vowels cannot easily be attributed to the vowel reduction associated with a casual speech style. Second, the increased perceptibility of tense vowels increases confidence in interpreting errors as problems with recall, rather than perception, of vowel targets.

As a result of including only tense vowels, the stimuli contained no weak syllables, by contrast with the typical English metrical stress pattern in which strong and weak syllables alternate. However, the lack of conformity to real words may be seen as an additional control for familiarity effects, further reducing the possibility that the correct vowel in any syllable could be guessed.

Table 1. Age and standard (Z -) score comparisons in children with (LI) and without (LN) language impairment.

	LI ($n = 20$)		LN ($n = 20$)		t	p
	M	SD	M	SD		
Age (in months)	93.70	10.59	94.00	10.85	0.16	.43
PPVT-R ^a	-1.85	0.98	0.06	1.17	5.60	<.001
TOLD-2 ^b	-1.59	0.62	-0.24	0.87	5.67	<.001
TONI 2 ^c	-0.77	1.05	-0.09	0.96	2.13	<.02

^aPeabody Picture Vocabulary Test–Revised (Dunn & Dunn, 1981).

^bSpoken Language Quotient, Test of Language Development–2 Primary (Newcomer & Hammill, 1988) or Test of Language Development–2 Intermediate (Hammill & Newcomer, 1988).

^cTest of Nonverbal Intelligence 2 (Brown, Sherbenou, & Johnsen, 1990).

Table 2. Phonetic transcriptions of the nonwords at each length.

One syllable	Two syllables	Three syllables	Four syllables
/naib/	/teivak/	/ʃɪnɔɪtaub/	/veɪtaɪfaɪdɔɪp/
/voʊp/	/ʃoʊvæg/	/naɪʃoʊveɪb/	/dævoʊnɔɪʃɪg/
/taʊdʒ/	/væʃaɪp/	/dɔɪtaʊvæb/	/naɪʃɔɪtaʊvʊb/
/dɔɪf/	/nɔɪtaʊf/	/teɪvɔɪʃaɪg/	/tævɔɪʃaɪg/

4. To reduce the predictability of consonant phonemes in the two possible syllable positions within the nonwords (onset or coda), consonants were assigned to occupy only those syllable positions in which they occurred $\leq 25\%$ of the time, according to data on the percentage of occurrence of each consonant in word-initial and word-final position (Shriberg & Kent, 1982, p. 429). Word-medial consonants were treated as syllable onsets and thus also had to occur $\leq 25\%$ of the time in word-initial position.

5. To ensure that accurate repetition of a nonword required that each of its phonemes be recalled independently, no consonants or vowels occurred more than once within a given nonword.

The four nonwords at each length were randomized to yield a consistent order of presentation progressing from the shortest, one-syllable, to the longest, four-syllable, nonwords. Nonwords were spoken by a trained adult female speaker wearing a head-mounted microphone (Shure SM10A) into an audio recorder (Marantz PMD 201); the speaker paused for approximately 3 s between each nonword. The speaker had previously practiced producing each nonword at a consistent rate, assigning primary stress to the second syllable of the four-syllable nonwords, and to the first syllable of all others. The duration of the entire task was 90 s.

The entire audiotape was subsequently transcribed independently by two research assistants who had no advance knowledge of the nonwords; point-by-point agreement for transcription of each phoneme was 100%, and corresponded exactly with the intended phonemes. In addition, transcribers agreed on the intended location of the syllable receiving primary stress in all of the multisyllabic nonwords.

The duration of each stimulus was measured using the Computerized Speech Laboratory Model 4300 B (Kay Elemetrics Corporation, 1994). The average duration of stimuli at each length was as follows: one-syllable = 622 ms; two-syllable = 918 ms; three-syllable = 1248 ms; four-syllable = 1504 ms, with standard deviations at each length ranging from 7 to 11 ms. This suggests that the speaker was successful in speaking each stimulus at a consistent rate, and that stimulus durations were within the range of expected values for citation-form productions of words containing one or more tense vowels.

Administration and Scoring

The nonword repetition task was administered to subjects individually as part of a larger test battery that included the norm-referenced tests reported in Table 1, among others, as well as a 10 min conversational language sample. All subjects passed a hearing screening

(ASHA, 1990) on the day of testing. All tasks were administered and scored by trained graduate research assistants who were blinded to subjects' intervention status. The nonword repetition task was presented under headphones at a comfortable listening level in a quiet location using a high quality cassette recorder (Marantz PMD 201). Subjects heard each nonword only once. The audiotaped instructions were: "Now I will say some made-up words. Say them after me exactly the way that I say them." The subjects' responses were audiorecorded by an external microphone onto a second portable cassette recorder for broad phonetic transcription.

Each phoneme (consonant or vowel) was scored as correct or incorrect in relation to its target phoneme. Phoneme substitutions and omissions were scored as incorrect; distortions of a phoneme were scored as correct. Phoneme additions were not counted as errors, because we were interested in the extent to which participants were able to represent the target phonemes in memory long enough to repeat them; additions by definition do not reflect a loss of information about the target phonemes themselves. Thus, although we have argued that additions might provide interesting evidence on the question of whether poor nonword repetition performance can be attributed to motoric limitations (Dollaghan, Biber, & Campbell, 1995), for purposes of the present report they were tallied separately and will not be discussed further.

In those cases in which a subject did not recreate the syllable structure of the nonword (adding or omitting one or more syllables), individual phoneme scoring proceeded after aligning the syllable sequence produced by the subject as nearly as possible to that of the target, using vowels repeated as syllable *anchors* to maximize the subject's score. For example, a subject who responded to the four-syllable target /dævounɔɪfɪg/ by saying /vɔɪfɪp/ would be scored as having attempted the second and fourth syllables, and having omitted the first and third; phoneme-by-phoneme scoring proceeded accordingly.

The number of phonemes repeated correctly was then divided by the total number of phoneme targets, resulting in a Percentage of Phonemes Correct (PPC) score at each nonword length (1PPC, 2PPC, 3PPC, 4PPC), and for the entire set of nonwords (TOTPPC).

Reliability

Audiotapes from eight randomly selected subjects (20%), four from each group, were transcribed independently by a second trained listener. Phoneme-by-phoneme percentages of agreement for judgments of correctness ranged from 91–99%, with an average of 94%.

Table 3. Percentage phonemes correct at each nonword length by children with (LI) and without (LN) language impairment.

Group	LN (n = 20)	LI (n = 20)
Nonword length		
1-syllable	91 (06)	86 (09)
2-syllable	92 (07)	83 (10)
3-syllable	90 (09)	68 (20)**
4-syllable	71 (11)	50 (16)**
Total	84 (07)	66 (12)**

** $p < .01$

Results and Discussion

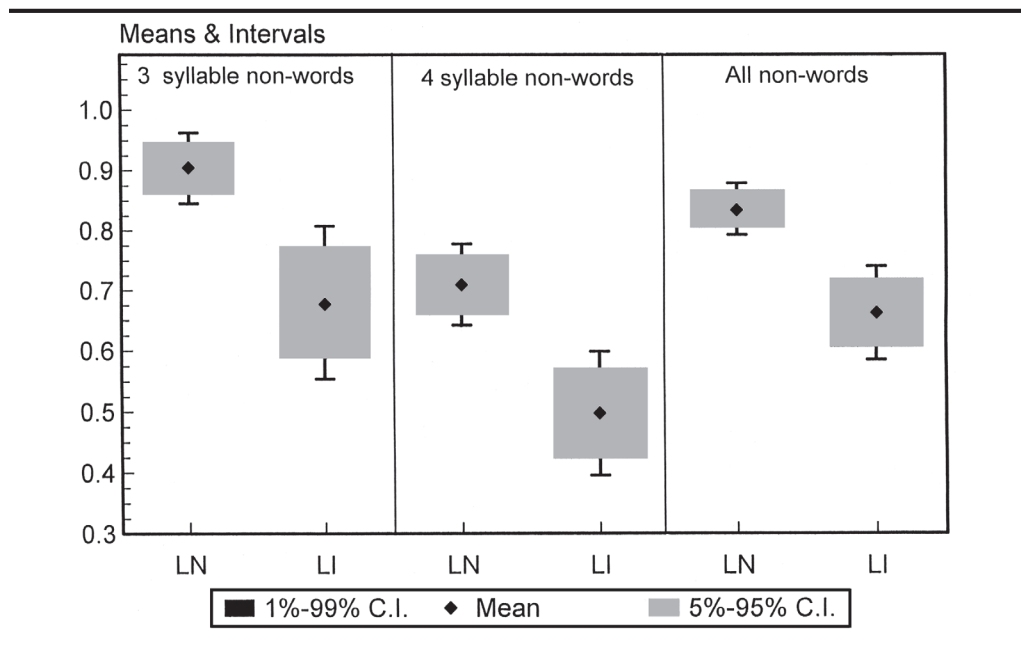
Means and standard deviations for Percentage Phonemes Correct at each nonword length, and for the task overall, appear in Table 3. A 2 [Group] × 5 [Length] ANOVA, with repeated measures on the second factor, revealed significant main effects of group [$F(1, 39) = 28.60, p < .01$] and length [$F(1, 4) = 80.93, p < .01$], and a significant Group × Length interaction ($F = 9.14, p < .01$). Scheffe tests with alpha set at $p < .01$ revealed that 3PPC, 4PPC and TOTPPC were significantly lower in the group with LI than in the group with LN. In both groups, 4PPC was significantly lower than PPCs at all shorter lengths. In the group with LI only, 3PPC was significantly lower than PPCs at the two shorter lengths. Figure 1 depicts the means and confidence intervals (95% and 99%) for the variables on which there were

significant group differences (3PPC, 4PPC, and TOTPPC); there was no overlap at the 99% confidence intervals of the two groups on these three nonword measures.

The well-known association between language and phonological impairments (e.g., Shriberg & Kwiatkowski, 1994) makes it important to ask whether the poor nonword repetition performance of the children in the LI group could be accounted for by constraints on their phonetic inventories. This issue has been addressed only indirectly in previous investigations, but the phoneme-by-phoneme scoring employed in the present investigation, coupled with the availability of information from other articulation and language testing, enabled a direct analysis. To do so, we examined repetitions of the 11 consonant phonemes on the nonword task by each child in the LI group. For each consonant repeated incorrectly at least once, we looked for evidence that it was in the child's phonetic inventory, as inferred from its production elsewhere on the nonword task, on subtests VII or IV of the TOLDP-2, or during the child's spontaneous language sample.

All of the consonants repeated incorrectly by these children during the nonword task were nonetheless in their phonetic inventories; in fact, 93% of the consonants on which errors occurred were produced correctly at some other point during the nonword task. For the remaining 7% of consonant errors, evidence that the target phonemes were in the child's inventory was readily obtained from the audio recorded TOLDP-2 or the language sample. This suggests that the poor

Figure 1. Means and confidence intervals (95% and 99%) for percentage phonemes correct for three-syllable (3PPC) and four-syllable (4PPC) nonwords, and for the entire nonword repetition task (TOTPPC) in the group with language impairment (LI) and the group developing language normally (LN).



nonword repetition performance of the children with LI cannot be explained by limitations in their consonant inventories.

Although none of the children with LI had been diagnosed as mentally retarded, and the problems of cognitive testing in such children are well known (e.g., Lahey, 1990), the fact that the average score of the children with LI on the TONI-R (Brown, Sherbenou, & Johnsen, 1990) was lower than that of the children with LN makes it important to ask whether poor nonword repetition performance could be attributed to poor nonverbal reasoning performance, possibly through a mediating influence of memory skills on both tasks. To address this question, we examined the association between TOTPPC and TONI-R score in the group with LI. The Spearman rank order correlation coefficient was .04 ($p > .85$), indicating that the group difference in nonword repetition cannot be attributed to the group difference in TONI-R performance.

Finally, although the ethnic distribution in the LI and LN groups was nearly identical, and Campbell et al. (1997) showed no evidence of bias in a similar nonword repetition task, the potential for dialect differences to influence nonword repetition performance made it important to ask whether the African American and White participants performed differently on this nonword repetition task. Of course, Black English Vernacular (BEV) or Inner City English (Akmajian, Demers, Farmer, & Harnish, 1995) is not associated exclusively with one ethnic group (Seymour & Seymour, 1981; Washington & Craig, 1992), nor do there appear to be clear criteria for determining whether a speaker uses BEV. However, some patterns associated with BEV, such as deletion or substitution of voiceless cognates for the nine nonwords that ended in voiced stops, could result in lowered scores by speakers of BEV. We had previously (Dollaghan & Trice, 1995) found no difference between ten pairs of age-matched African American and White children, with and without language impairment, at any nonword length on the present task, even though African American participants produced significantly more syntactic forms associated with Black English Vernacular in conversation than did white participants. Similarly, in the present study, the TOTPPC of the 25 African American participants ($M = 75\%$) and of the 9 White participants ($M = 74\%$) did not differ significantly (Wilcoxon Rank Sum $z = .29$; $p = .77$).

The results from Study 1 thus corroborate previous evidence that children with language impairments repeat nonwords less accurately than do their peers, and suggest that a disparity in language knowledge cannot explain this group difference. Further, the lack of overlap in scores for the three- and four-syllable nonwords and for the task overall suggests that this measure might

be an efficient and effective means of distinguishing between children with and without language impairments within this age range. In Study 2 we examined this possibility.

Study 2

Purpose

As noted earlier, children with LI perform less well than children with LN on a wide variety of tasks, but there is typically a great deal of overlap between the distributions of these groups. Thus, the mere fact that the groups differ significantly on some measure enables no inferences about the extent to which the scores of individual children on the measure would accurately identify them as LI or LN. The fact that the distributions of total, three- and four-syllable PPCs in nonword repetition were so distinct in Study 1 invited a further exploration of the extent to which these measures might differentiate individual children with LI or LN. We recently argued that nonword repetition is among a set of processing-dependent measures that are better suited to identifying fundamental language processing deficits than typical language tests, because they are less dependent on the test-taker's prior knowledge and experience (Campbell, et al., 1997). A finding that performance on the brief, easily administered nonword repetition task successfully distinguished children with LI from children with LN would contribute to efforts to develop unbiased, reliable, and efficient ways to identify children whose language impairments cannot be attributed to experiential factors.

The purpose of Study 2, accordingly, was to compare the clinical utility of the nonword repetition task and a knowledge-dependent, norm-referenced language measure in distinguishing between school-age children with and without language impairment, again using enrollment in language intervention, as determined independently by ASHA-certified public school speech-language pathologists, as our gold standard for the condition of language impairment. The relative clinical utility of the two measures was determined by calculating likelihood ratios (Sackett et al., 1991) for several levels of performance on each task in a larger ($N = 85$) group of children with and without LI.

As described by Sackett et al., "...a likelihood ratio expresses the odds that a given level of a diagnostic test result would be expected in a patient with (as opposed to one without) the target disorder" (p. 120). These authors noted that, by contrast with the more familiar constructs of sensitivity and specificity of a diagnostic test,

likelihood ratios are less sensitive to changes in the pretest probability of a disorder. In addition, likelihood ratios can be calculated for several levels of performance on a diagnostic test, and thus the performance of a given child can be linked to its likelihood ratio to yield the odds that he or she is a member of the group with LI or LN. Tests that have high likelihood ratios for ruling a patient into a group having some condition, and low likelihood ratios for ruling an unaffected individual out of the clinical group, thus offer the diagnostician powerful evidence concerning an individual patient's status with respect to a disorder. As described by Sackett et al., likelihood ratios of 20 or more for a positive result on a diagnostic test can be described as "high," because they yield posttest probabilities of 95% or more that the target disorder is present. At the other end of the performance scale, for ruling out the presence of a target disorder, likelihood ratios close to zero for a negative result on a diagnostic test are also very informative. For example, a likelihood ratio of .08 for a negative test result would correspond to a posttest probability of the target disorder of less than 4%. Likelihood ratios between these two extremes are less informative. For example, a test result with a likelihood ratio close to 1.0 provides the clinician with little information beyond that which he or she had before giving the test.

We hypothesized that nonword repetition, as a processing-dependent measure designed to tap fundamental psycholinguistic processing operations, might better predict enrollment in language intervention than a knowledge-dependent measure of spoken language skill on which performance could be influenced by factors such as language background, experiential history, and vocabulary knowledge. We surmised that clinicians' judgments about children's need for language intervention, based on multiple sources of evidence interpreted within the larger context of the school and community, would be better reflected in the processing-dependent than in the knowledge-dependent language measure. Thus, we predicted that the nonword repetition task would yield more informative likelihood ratios than the knowledge-dependent language test. If clinicians' judgments about the need for language intervention had been based largely or entirely on knowledge-dependent test performance, likelihood ratios for the knowledge-based test would predict group membership with a high degree of accuracy.

Method

Participants

Participants in Study 2 were the 40 age-matched children from Study 1, in addition to the remaining 45 school-age children from the larger study whose ages did not allow them to be matched into LI-LN pairs. The

sample for Study 2 thus included 85 children, ranging in age from 5;8 to 12;2, who had been referred by a certified speech-language pathologist as either enrolled in language intervention (LI; $n = 44$) or developing language normally and not enrolled in speech or language services (LN; $n = 41$). The present sample of 85 included all subjects thus identified whose parents had responded affirmatively to a mailing soliciting participation, who had English as their first language, who passed a hearing screening (ASHA, 1990) on the day of testing, and who did not exhibit fluency disorders that would be expected to interfere with nonword repetition performance. Eight children were not included in the final sample of 85 because 5 failed the hearing screening, 2 exhibited fluency disorders, and 1 had a first language other than English. In addition, 1 subject referred to the study was excluded when he was unable to complete the norm-referenced testing, apparently due to an inability to understand any of the testing tasks.

Of the remaining 85 participants, 66% were male, and parent-reported ethnicity was as follows: 58% African American, 34% White, 2% Hispanic, and 5% mixed. One parent declined to provide information on her child's ethnicity. The ethnic distribution of this sample corresponded well to the ethnic distribution of the urban school district that all but 4 participants attended. The distributions of gender and parent-reported ethnicities within the groups designated as LI and LN were similar, i.e., 61% of the LI group and 54% of the LN group were African American; 70% of the LI group and 61% of the LN group were male.

Socioeconomic status of participants was not assessed formally, but caregivers completed a questionnaire concerning the number of parents living at home, their occupations, and the level of education they had completed. By these reports, the majority of children in both the LI and LN groups (62% and 56%, respectively) were living in single parent homes; the majority of parents, again in both groups (LI = 65%; LN = 56%), were either unemployed or held jobs involving unskilled labor (e.g., food service worker, child care worker, truck driver, manual laborer). Thus, it is reasonable to infer that the majority of the subjects in both groups were from lower-income households. With respect to highest parent educational level, in both groups the majority of parents had completed at least some college (LI = 61%; LN = 68%); 7% of the LI parents and 3% of the LN parents had not completed high school.

Procedure

The nonword repetition task was administered to all participants as part of the battery of hearing, nonverbal reasoning, language, and speech tests described in

Study 1; the Test of Language Development–2 (TOLD-2; Hammill & Newcomer, 1988; Newcomer & Hammill, 1988) was administered as part of this battery. From the nonword repetition task, percentage of phonemes correct (PPC) was calculated for each nonword length and for the task overall, using procedures identical to those in Study 1. From the TOLD-2, Spoken Language Quotients (SLQs) were calculated according to the procedures defined by the TOLD-2 manual.

To examine the internal consistency of the nonword repetition task, the correlation of each subject's TOTPPC on the odd and even nonwords was calculated.

To calculate likelihood ratios for the nonword task, the frequency distributions of PPC for the nonword task as a whole (TOTPPC), for the three-syllable nonwords (3PPC), and for the four-syllable nonwords (4PPC) were then examined according to the methods described by Sackett et al. (1991) to determine the levels of performance that best differentiated the children with LI and LN. Similarly, the distributions of standard scores (z -scores) on the TOLD-2 SLQ were examined to determine the performance levels that best differentiated between the groups. Likelihood ratios were then calculated for each measure as described below.

Reliability

Audiotapes from 30% of the subjects in each group (LI and LN) were transcribed independently by a second trained listener. Phoneme-by-phoneme percentages of agreement for judgments of correctness ranged from 90–100%; the average percentage of agreement in both groups was 95%.

Results and Discussion

The split-half (odd-even) reliability for the nonword repetition task was high ($r = .849$, $p < .01$), suggesting that the 16 nonword stimuli meet accepted standards for internal consistency. Table 4 illustrates the process by which likelihood ratios are calculated using PPCs obtained by the subjects with LI and LN on the entire

nonword task (TOTPPC). The likelihood ratio for a positive test result (to rule in the presence of the target disorder), defined as a TOTPPC of 70% or lower, is calculated by dividing the true positive rate (the number of children with LI with TOTPPCs at or below 70%, which is 27/44 or 0.6136) by the false positive rate (the number of children with LN with TOTPPCs at or below 70%, which is 1/41 or .0244). The resulting likelihood ratio (0.6136/0.0244) for ruling a child into the LI group based on a TOTPPC of 70% or lower was 25.15; this means that a TOTPPC of 70% or lower on this nonword repetition task was 25 times more likely to come from a child with LI than from a child with LN. As described by Sackett et al. (1991), a likelihood ratio of 25.15 in a sample with approximately equal numbers of affected and unaffected participants (as in the present sample) corresponds to a posttest probability of LI of more than 95%; in their terms, a total PPC of 70% or lower would be adequate to “rule in” the presence of a language disorder in a school-age child with a very small likelihood of error.

On the other end of the spectrum, the likelihood ratio for a negative test result sufficient to rule out the presence of the target disorder is calculated for a TOTPPC of 81% or greater by dividing the false negative rate (the number of children with LI who had TOTPPCs this high, which is 1/44 or .0227) by the true negative rate (the number of children with LN who had TOTPPCs this high, which is 28/41, or .6829). The resulting likelihood ratio for a negative test result (a TOTPPC of 81% or greater) is 0.03, which means that a TOTPPC this high was less than one-twentieth as likely to come from a child with LI as from a child with LN. Thus, a child who obtains a total PPC of 81% or more could be ruled out of the LI group with a high degree of confidence.

Likelihood ratios for levels of TOTPPC performance between these two extremes were also calculated. As shown in Table 4, TOTPPCs from 71–74% corresponded to an “intermediate high” (Sackett et al., 1991) likelihood ratio of 3.11, meaning that TOTPPCs in this range are three times more likely to come from children with LI than LN. Additional diagnostic testing would be necessary in order to classify children with TOTPPCs in

Table 4. Likelihood ratios for total percentage phonemes correct (TOTPPC) in children with impaired and normal language.

	Language impaired ($n = 44$)		Language normal ($n = 41$)		Likelihood ratio
	Number	Proportion	Number	Proportion	
≤70	27	.6136	1	.0244	.6136/.0244 = 25.15
71–74	10	.2273	3	.0732	.2273/.0732 = 3.11
75–80	6	.1364	9	.2195	.1364/.2195 = 0.62
≥81	1	.0227	28	.6829	.0227/.6829 = 0.03

Table 5. Likelihood ratios for percentage phonemes correct from three-syllable (3PPC) and four-syllable (4PPC) nonwords.

	Language impaired (n = 44)		Language normal (n = 41)		Likelihood ratio
	Number	Proportion	Number	Proportion	
3PPC					
≤63	20	.4545	1	.0244	.4545/.0244 = 18.63
64–78	11	.2500	2	.1951	.2500/.1951 = 5.12
79–89	11	.2500	12	.2927	.2500/.2927 = 0.85
≥90	2	.0454	26	.6341	.0454/.6341 = 0.07
4PPC					
≤46	23	.5227	2	.0488	.5227/.0488 = 10.71
47–63	13	.2955	4	.0976	.2955/.0976 = 3.03
64–74	7	.1591	16	.3902	.1591/.3902 = 0.41
≥75	1	.0227	19	.4634	.0227/.4634 = 0.05

this range correctly. The likelihood ratio for TOTPPCs between 75–80%, 0.62, is intermediate or indeterminate. Additional evidence would be needed to classify children with TOTPPCs in this range.

TOTPPCs from this sample of school-age children generated the most informative likelihood ratios, but levels of performance on three- and four-syllable nonwords alone also yielded very informative likelihood ratios (see Table 5). Although TOTPPC enabled the most accurate classification of children with LI or LN, levels of 3PPC and 4PPC also distinguished between children with LI and LN with a high degree of confidence.

Finally, Table 6 compares the likelihood ratios for TOTPPC with those generated by various levels of performance (*z*-scores) from the Spoken Language Quotient (SLQ) of the TOLD-2. The highest likelihood ratio that could be generated for a positive result on SLQ, 3.73, was associated with an SLQ 1.5 *SDs* or more below the mean for the child's age. This is an "intermediate high" (Sackett et al., 1991) likelihood ratio, corresponding to a posttest probability of LI status of somewhat less than 70%; this means that an SLQ ≥ -1.5 *SDs* would not be sufficient to classify a child as LI without additional testing. Similarly, the best likelihood ratio for ruling out the presence of a language impairment based on SLQ was 0.91; the proximity of this likelihood ratio to 1.0 indicates that SLQ performance offered virtually no information concerning whether a child could be ruled out of the group with LI.

It is important to note that the TOLD-2 was designed for objectives other than the rapid identification of children with LI, and there is a considerable body of evidence on the lack of congruence between norm-referenced test results and clinical judgments about child language impairment (Aram et al., 1993; Dunn et al., 1996; Stark & Tallal, 1981). However, the data in Table 6 suggest that

when the diagnostic objective is screening, (i.e., identifying school-age children who are in need of more comprehensive language testing) a great deal of information is provided by the nonword repetition task, which requires just 90 s to administer and approximately 6 min to transcribe and score. The processing-dependent measure of nonword repetition performance yielded much more accurate information concerning a child's language intervention status in much less time than a knowledge-dependent test.

General Discussion

In these studies, we found that children with and without language impairments performed quite differently on a nonword repetition task that had been carefully constructed to minimize its familiarity and predictability and to maximize scoring accuracy and reliability. Further, we demonstrated that certain levels of nonword repetition performance were extremely powerful predictors of language status, differentiating

Table 6. Likelihood ratios for total percentage phonemes correct (PPC) and spoken language quotient^a, for school-age children in predicting language impairment.

Total PPC		SLQ <i>z</i> -score	
Level	Likelihood ratio	Level	Likelihood ratio
≤70%	25.15	≥-1.5	3.73
71–74%	3.11	0 to -1.49	0.64
75–80%	0.62	0.1 to 0.49	0.47
≥81%	0.03	≥0.5	0.91

^aTest of Language Development-2 (Hammill & Newcomer, 1988; Newcomer & Hammill, 1988)

children enrolled in language intervention from children developing language normally almost perfectly. Finally, we demonstrated that certain levels of nonword repetition performance were far superior to levels of performance on a knowledge-dependent language test in identifying children who had been independently diagnosed and enrolled in language intervention in a large, urban public school district. A substantial percentage of the participants were reported by their parents to be African American and the majority of were from lower-income families. Our findings are consistent with previous evidence (Campbell et al., 1997) that processing-dependent measures minimize the problem of test bias associated with different income and educational levels (e.g., Brooks-Gunn, Klebanov, & Duncan, 1996; Hart & Risley, 1995). The present results augur well for efforts to develop measures that provide diagnostically meaningful information for children from a variety of ethnic, educational, and socioeconomic backgrounds.

The results also illustrate a more general strategy for evaluating the clinical utility of measures that could substantially increase our diagnostic acumen. If likelihood ratios were available for every measure of language that is believed to differentiate children with and without LI, assessment time could be devoted solely to the most informative measures. Likelihood ratios can also be determined for sequences of measures, when results from an initial test yield performance levels in the questionable or indeterminate range (Sackett et al., 1991).

Determining likelihood ratios requires an accepted gold standard for the presence of a language disorder. Nearly every published study of children with LI devotes significant space to describing criteria for defining LI. The necessity for such individualized definitions is usually attributed to the performance heterogeneity of children with LI. However, the heterogeneity may result from the measures, not the children. Likelihood ratios offer clear grounds for deciding whether a measure is sufficiently informative to warrant its use for identifying children with LI. In our data, the nonword repetition task yielded likelihood ratios that were substantially more consonant with independent clinical judgments than those from a widely used knowledge-dependent language measure. However, the definitive evidence concerning accuracy of diagnosis of an impairment in any behavioral domain must come from studies documenting differences in short- and long-term clinical outcomes.

Children with language impairments have notable deficits in nonword repetition that cannot be attributed to differences in their language knowledge. Although these findings do not pinpoint the locus of these deficits, efforts toward this end continue (e.g., Edwards & Lahey, 1996; Gathercole & Martin, 1996; van der Lely & Howard, 1993). Successful performance of the

nonword task requires several processing operations that are assumed to be involved in language learning, including transforming the acoustic-phonetic sequence into its constituent phonemes, maintaining the ordered and phonologically coded string in working memory, and organizing the articulatory output. Deficits in any or all of these operations could have negative consequences for nonword repetition and language learning tasks (e.g., creating new lexical entries and formulating sentences). The extent to which nonword repetition performance distinguished between children enrolled in language intervention and children developing language normally strengthens the rationale for continuing to explore the processing variables that underlie both nonword repetition and more general language performance.

The question of whether nonword repetition performance can also identify younger children with clinically significant language impairments is being examined in large ($N > 100$) samples of ethnically and socioeconomically diverse children between ages of 3;0–3;3, and 4;0–4;3 (Paradise et al., 1997). These data will enable us to determine whether the nonword repetition task can identify young children whose language deficits do not reflect experiential differences. We are also examining the utility of this measure to distinguish adults with and without histories of language intervention during childhood, a prerequisite to establishing a phenotype for behavioral and molecular genetic studies of child language impairment.

Acknowledgments

This study was supported by a research grant from the National Institute on Deafness and Other Communication Disorders (R29DC01858) of the National Institutes of Health. Thanks to Denise Balason, Christine Gable, Tara Jackson, Andrea Schwartz, and Caroline Trice for their help in data collection and analysis. Thanks to Ellen Estomin and Barbara Leonard of the Pittsburgh Public Schools for their help in recruiting subjects.

References

- Akmajian, A., Demers, R. A., Farmer, A. K., & Harnish, R. M.** (1995). *Linguistics*. Cambridge, MA: MIT Press.
- American Speech-Language-Hearing Association** (1990). Guidelines for screening for hearing impairment and middle-ear disorders. *Asha*, 32 (Suppl. 2), 17–24.
- Aram, D. M., Morris, R., & Hall, N. E.** (1993). Clinical and research congruence in identifying children with specific language impairment. *Journal of Speech and Hearing Research*, 36, 580–591.
- 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*, 36, 1–13.
- Brooks-Gunn, J., Klebanov, P. K., & Duncan, G. J.** (1996).

- Ethnic differences in children's intelligence test scores: Role of economic deprivation, home environment, and maternal characteristics. *Child Development*, 67, 396–408.
- Brown, L., Sherbenou, R., & Johnsen, S.** (1990). *Test of Nonverbal Intelligence 2*. Austin, TX: Pro-Ed.
- Campbell, T. F., 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.
- Dollaghan, C., & Trice, C.** (1995, November). *An evaluation of bias in a measure of phonological memory*. Poster session presented at the annual meeting of the American Speech-Language-Hearing Association, Orlando, FL.
- Dollaghan, C., Biber, M., & Campbell, T.** (1993). Constituent syllable effects in a nonsense-word repetition task. *Journal of Speech and Hearing Research*, 36, 1051–1054.
- Dollaghan, C., Biber, M., & Campbell, T.** (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics*, 16, 211–222.
- Dunn, L., & Dunn, L.** (1981). *Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance Service.
- Dunn, M., Flax, J., Sliwinski, M., & Aram, D.** (1996). The use of spontaneous language measures as criteria for identifying children with specific language impairment: An attempt to reconcile clinical and research incongruence. *Journal of Speech and Hearing Research*, 39, 643–654.
- Edwards, J., & Lahey, M.** (1996, June). *Why do children with specific language impairment repeat nonwords less accurately than their peers?* Poster session presented at the 17th Symposium on Research in Child Language Disorders, University of Wisconsin–Madison.
- Fazio, B. B., Naremore, R. C., & Connell, P. J.** (1996). Tracking children from poverty at risk for specific language impairment: A 3-year longitudinal study. *Journal of Speech and Hearing Research*, 39, 611–624.
- Fischer, C. S., Hout, M., Sanchez Jankowski, M., Lucas, S. R., Swidler, A., & Voss, K.** (1996). *Inequality by design*. Princeton, NJ: Princeton University Press.
- Gathercole, S. E., & Baddeley, A. D.** (1990a). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, 29, 336–360.
- Gathercole, S. E., & Baddeley, A. D.** (1993). *Working memory and language*. Hove, UK: Lawrence Erlbaum Associates.
- Gathercole, S. E., & Martin, A. J.** (1996). Interactive processes in phonological memory. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 73–100). Hove, UK: Psychology Press.
- Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D.** (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, 12, 349–367.
- Hammill, D., & Newcomer, P.** (1988). *Test of Language Development 2 Intermediate*. Austin, TX: Pro-Ed.
- Hart, B., & Risley, T. R.** (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes.
- Kay Elemetrics Corporation** (1994). *CSL: Computerized Speech Laboratory, Model 4300B*. Pine Brook, NJ.
- Lahey, M.** (1990). Who shall be called language disordered? Some reflections and one perspective. *Journal of Speech and Hearing Disorders*, 55, 612–620.
- Leonard, L.** (1987). Is specific language impairment a useful construct? In S. Rosenberg, (Ed.), *Advances in applied psycholinguistics, Vol. 1* (pp. 1–39). New York: Cambridge University Press.
- Montgomery, J. W.** (1995). Sentence comprehension in children with SLI: The role of phonological working memory. *Journal of Speech and Hearing Research*, 38, 187–199.
- Newcomer, P., & Hammill, D.** (1988). *Test of Language Development 2 Primary*. Austin, TX: Pro-Ed.
- Paradise, J. P., Rockette, H. E., Colborn, D. K., Bernard, B. S., Smith, C. G., Kurs-Lasky, M., & Janosky, J. E.** (1997). Otitis media in 2253 Pittsburgh-area infants: Prevalence and risk factors during the first two years of life. *Pediatrics*, 99, 318–333.
- Paul, R.** (1995). *Language disorders from infancy through adolescence*. St. Louis, MO: Mosby.
- Records, N. L., & Tomblin, J. B.** (1994). Clinical decision making: Describing the decision rules of practicing speech-language pathologists. *Journal of Speech and Hearing Research*, 37, 144–156.
- Sackett, D. L., Haynes, R. B., Guyatt, G. H., & Tugwell, P.** (1991). *Clinical Epidemiology*. Boston: Little, Brown.
- Seymour, H. N., & Seymour, C. M.** (1981). Black English and standard American English contrasts in consonantal development of four- and five-year-old children. *Journal of Speech and Hearing Disorders*, 46, 274–280.
- Shriberg, L. D., & Kent, R. D.** (1982). *Clinical phonetics*. New York: Wiley.
- Shriberg, L. D., & Kwiatkowski, J.** (1994). Developmental phonological disorders I: A clinical profile. *Journal of Speech and Hearing Research*, 37, 1100–1126.
- Snowling, M., 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., & Tallal, P.** (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders*, 46, 114–122.
- Tomblin, J. B., Records, N. L., & Zhang, X.** (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research*, 39, 1284–1294.
- 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.
- Washington, J., & Craig, H.** (1992). Articulation test performances of low-income, African-American pre-schoolers with communication impairments. *Language, Speech and Hearing Services in Schools*, 23, 203–207.

Received July 21, 1997

Accepted February 27, 1998

Contact author: Chris Dollaghan, University of Pittsburgh, Department of Communication Science and Disorders, 4033 Forbes Tower, Pittsburgh, PA 15260.
Email: dollagha+@pitt.edu