Reading Achievement Growth in Children With Language Impairments

Hugh W. Catts
Mindy Sittner Bridges
Todd D. Little
University of Kansas, Lawrence

J. Bruce Tomblin
University of Iowa, Iowa City

Purpose: This study examined the reading achievement growth of children with language impairments (LI) across the school grades. The authors sought to determine whether children with LI demonstrate a delayed, deficit, or cumulative pattern of reading achievement growth when compared with children with typical language (TL).
Method: A group of 225 children with LI and a group of 379 children with TL were identified in kindergarten and were administered multiple measures of word recognition and reading comprehension in 2nd, 4th, 8th, and 10th grades.
Results: Confirmatory factor analyses indicated that the constructs of word recognition and reading comprehension were invariant across grades and groups of children with LI and TL. Further analyses indicated that a multiple group latent growth curve analysis was appropriate. This analysis showed that children with LI differed significantly from children with TL in initial level (2nd grade) of word recognition and reading comprehension, but they did not differ significantly in the shape of their growth trajectories.
Conclusion: These results are consistent with a deficit model of reading growth in children with LI. Findings are discussed in terms of implications for early identification.

KEY WORDS: language impairment, reading growth, reading achievement

It is well recognized that oral language development lays the foundation for reading achievement (Catts, Fey, Zhang, & Tomblin, 1999; NICHD Early Child Care Research Network, 2005; Storch & Whitehurst, 2002). Children with typical language (TL) development generally demonstrate normal reading achievement, whereas those with spoken language impairments (LI) often experience significant problems learning to read. Over the last 20 years, numerous studies have documented the poor reading outcomes of children with LI (Bishop & Adams, 1990; Botting, Simkin, & Conti-Ramsden, 2006; Catts, 1993; Catts, Fey, Tomblin, & Zhang, 2002; Menyuk et al., 1991; Naucler & Magnusson, 1998; Roth, Cooper, & Speece, 2002; Silva, McGee, & Williams, 1987; Tallal, Curtis, & Kaplan, 1989). Most of these studies have focused on early reading outcomes and have revealed the difficulties that these children have acquiring basic literacy skills. Few investigations, however, have systematically followed children with LI into middle and high school (but see Snowling, Bishop, & Stothard, 2000). As a result, very little information is available concerning the growth of reading achievement across the school grades in this population.

The reading achievement growth trajectories of children with LI could take several forms when compared with those of children with TL development. Children with LI could show lower initial reading achievement but parallel growth over the school years. This has been referred to as a deficit model of reading growth (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996). Alternatively, children with LI could demonstrate lower initial reading achievement but accelerated growth, which would allow them to catch up to their TL peers during the later school years. This pattern is known as a *developmental lag* or *delayed pattern of reading growth* (Francis et al., 1996). Finally, children with LI could get off to a slow start in reading achievement as well as demonstrate slower growth. This would lead to a widening of the gap in reading achievement between them and their TL peers. The latter pattern is characteristic of what some have called the *Matthew effect* (Stanovich, 1986) and of what others have referred to as a *cumulative trajectory model of achievement* (Leppänen, Niemi, Aunola, & Nurmi, 2004).

Variability in reading growth has been studied within the general population for some time. In a seminal article, Juel (1988) identified good and poor readers in first grade and compared their reading outcomes in fourth grade. She found that good readers typically remained good readers and that poor readers seldom "caught up" in reading achievement. Such results are supportive of either a deficit or cumulative trajectory pattern. Others have reported specific evidence of a cumulative trajectory pattern in which poor readers have continued to fall further and further behind good readers (Bast & Reitsma, 1998; Hayes & Grether, 1983; McKinney & Feagans, 1984; Williamson, Appelbaum, & Epanchin, 1991). For example, Bast and Reitsma (1998) reported an increase in variability in children's word reading scores in kindergarten through second grade but not through third grade. This increase in variability is indicative of the spreading out of growth trajectories (i.e., fan effect) associated with the Matthew effect.

The most comprehensive examination of the variability in reading growth in the general population was conducted in the Connecticut Longitudinal Study (Francis et al., 1996; Shaywitz et al., 1995). In an initial study, Shaywitz et al. (1995) followed approximately 400 children from first through sixth grade and reported no evidence of a cumulative growth trajectory pattern among poor readers. In a follow-up study, Francis et al. (1996) used individual growth curve analysis to test a deficit versus a developmental lag model of reading growth. They examined reading achievement from first to ninth grades in normal readers and two groups of poor readers from the sample. Results showed that all three groups displayed a quadratic growth pattern characterized by sharp initial growth through sixth grade and a plateau thereafter. Consistent with a deficit model, poor readers differed significantly from normal readers in initial level of reading achievement but did not differ significantly in the rate of development (i.e., shape of growth curve). Others have also reported no evidence of widening of the achievement gap between good and poor readers (Baker, Decker, & DeFries, 1984; Scarborough & Parker, 2003).

Although information about the long-term reading achievement of children with LI is limited, there are a few relevant studies. Beitchman, Nair, Clegg, and Patel (1986) identified a large group of children with LI ($N = 142$) from an epidemiologic sample of 5-year-old children. Follow-up testing of their academic outcomes was conducted at 12 and 19 years of age (Johnson et al., 1999; Young et al., 2002). The researchers reported that children with LI performed significantly less well in reading achievement than normal controls at each age. However, no direct comparisons of growth in reading were made. In another longitudinal study, Bishop, Snowling, and colleagues identified a convenience sample of children with LI at 4 years of age ($N = 87$) and conducted follow-up testing when the children were 5, 8, and 15 years of age (Bishop & Adams, 1990; Snowling et al., 2000). Reading achievement was assessed at the latter two time points with different measures. However, the researchers made comparisons between ages on the basis of standard scores. Although not optimal, these data provide some suggestions concerning reading growth. Snowling et al. (2000) reported that children with LI, specifically those with specific LI, showed a decline in reading achievement relative to normal controls between 8 and 15 years of age. This apparent cumulative growth pattern was particularly the case for word recognition in children whose performance IQ was less than 100.

Whereas the above studies provide some initial information about reading achievement growth in children with LI, they were not optimally designed to measure growth across the school years. In the present investigation, we used data from the Child Language Research Study (Tomblin, 1995) to directly examine the reading achievement growth trajectories from 2nd to 10th grade of children with LI. In this study, children with LI and children with TL were identified in kindergarten, and their reading achievement was assessed with multiple measures of word recognition and reading comprehension in 2nd, 4th, 8th, and 10th grades. Latent growth curve (LGC) analysis was used to compare the reading achievement growth trajectories of these groups of children.

Method

Participants

The participants in this study were a subsample of children who originally took part in an epidemiologic study of kindergarten children with LI (Tomblin et al., 1997). The epidemiologic investigation utilized a stratified cluster sample of 7,218 children. This normative sample was stratified by residential setting (i.e., rural, urban, suburban) and was cluster sampled by school building. All available kindergarten children in selected schools were screened for LI. Children who failed the screening, and a random sample who passed, were given a test battery of language and other measures. The results of this assessment were used to estimate the prevalence of kindergarten children with LI (Tomblin et al., 1997).

Upon completion of the epidemiologic study, 604 children who had received the test battery in kindergarten participated in a follow-up longitudinal investigation conducted by the Child Language Research Center (Tomblin, 1995). These children served as the participants for the present study. Complete data were available on 493 of these children through 10th grade. Expectation-maximization (EM) imputation was used to estimate the missing data to allow for the analysis of the full sample of 604 children (Graham & Schafer, 2002). All children in this study were monolingual English speakers and had no history of sensory deficits or neurological disorders. For further details concerning the participant sample, see Catts et al. (2002).

The results of the diagnostic test battery administered in kindergarten were used to identify participants with and without LI. Our definition of LI was consistent with that used in the epidemiologic study. This definition is based on a model of language that includes three domains of language (vocabulary, grammar, and narration) and two modalities (receptive and expressive). For this definition, a composite score was calculated for each domain and modality of language (see the *Materials* section). Participants were defined as having LI if their performance on at least two of five language composite scores was 1.25 standard deviations or more below the mean. These procedures identified 225 children with LI and 379 with TL. Included in this sample were 102 children with LI and 103 children with TL who had nonverbal cognitive abilities more than 1 standard deviation below the mean. To maintain adequate numbers in the LI and TL groups for the present analyses, we did not subdivide groups on the basis of nonverbal cognitive ability. However, we did use nonverbal cognitive ability as a covariate in the growth analyses.

Materials

Kindergarten Language Status

As part of the epidemiologic study, five subtests of the Test of Language Development–2: Primary (TOLD-2:P; Newcomer & Hammill, 1988) and a narrative story task (Culatta, Page, & Ellis, 1983) were utilized. Local norms were used to convert raw scores to standard scores. The standard scores, which were in the form of z scores, from the TOLD-2:P Picture Identification and Oral Vocabulary subtests were combined to form a vocabulary composite score. We used z scores from the TOLD-2:P Grammatic Understanding, Grammatic Completion, and Sentence Imitation subtests to form a grammar composite score, whereas we employed scores from the Narrative Comprehension and Recall measures (Culatta et al., 1983) as a narrative composite score. To derive a receptive language composite score, z scores from the Picture Identification, Grammatic Understanding, and Narrative Comprehension subtests were combined. Finally, to obtain an expressive language composite score, z scores from the Oral Vocabulary, Grammatic Completion, Sentence Imitation, and Narrative Recall were used. An overall language composite score was also calculated using the expressive and receptive language composite scores. As noted previously, children were identified as having LI if they performed at least 1.25 standard deviations below the mean on at least two of the five composite scores (Tomblin, Records, & Zhang, 1996). This corresponds to approximately at least 1.14 standard deviations below the mean on the overall language composite.

Nonverbal Cognitive Abilities

Nonverbal cognitive abilities were measured in kindergarten by the Block Design and Picture Completion subtests of the Wechsler Preschool and Primary Scale of Intelligence–Revised (Wechsler, 1989). These subtests measure a range of nonverbal cognitive abilities, including visual attention, visual recognition, visual–motor coordination, and spatial reasoning (Kaufman, 1979). Performance on the two subtests was combined to form a composite measure of nonverbal cognitive abilities.

Reading Achievement

Measures of word recognition and reading comprehension were administered in 2nd, 4th, 8th, and 10th grades.

Word recognition. Two subtests from the Woodcock Reading Mastery Tests–Revised (WRMT-R; Woodcock, 1987) and a measure from the Gray Oral Reading Test–Third Edition (GORT-3; Wiederholt, & Bryant, 1992) were used to assess word recognition. The Word Identification subtest from the WRMT-R measured participants' ability to accurately pronounce printed English words ranging from high to low frequency of occurrence. The Word Attack subtest assessed participants' ability to read pronounceable nonwords varying in complexity. For both measures, raw scores were converted to W scores, which are Rausch-based scores on an interval scale with a constant metric. A score of 500 corresponds to the average performance level of a student at the beginning of fifth grade. The adherence to an interval scale with a constant metric makes W scores well suited for examining reading growth. The GORT-3 included a text reading accuracy measure that was used in our index of word recognition. In this task, participants read aloud short passages, and word reading accuracy was assessed.

Reading comprehension. The Passage Comprehension subtest of the WRMT-R and the comprehension standard score from the GORT-3 were used to assess reading comprehension across grades. The WRMT-R subtest uses a cloze procedure, in which participants read short passages aloud and provide a missing word. For the GORT-3, participants read aloud a short passage and answered multiple-choice questions concerning the passage. In second and fourth grades, participants were administered the Reading Comprehension subtest of the Diagnostic Achievement Battery–2 (DAB-2; Newcomer, 1990). This subtest was not appropriate for 8th and 10th grades and was replaced by the reading comprehension component of the Qualitative Reading Inventory–Second Edition (QRI-2; Leslie & Caldwell, 1995). For both of the latter measures, participants read short passages and answered open-ended questions. Because these latter two measures were not given at all four grades, they could not be used in the latent growth model. However, they were considered in the confirmatory factor analysis (CFA).

Procedure

A total of 14 examiners participated in the administration of the test battery in kindergarten and another three in the administration of the test batteries in the other grades. Seven of the examiners were certified speech-language pathologists, and the remaining had undergraduate degrees in speech and hearing (n = 3) or education (n = 7). All language measures were administered by examiners certified in speech-language pathology. In addition, all examiners received approximately 1 week of training by the investigators on the administration of the testing protocols. Testing was conducted in specially designed vans parked at the participants' schools or homes.

Missing Data and Analyses

We used SAS Institute's (www.sas.com) Proc MI with MCMC estimation (EM initial) to impute the 6.62% of data that were missing from the 604 participants. For this estimation, we included all variables available in the data set to inform the missing data as richly as possible under the assumption that data were missing at random. After imputation, we standardized the constructs of word recognition and reading comprehension to have a common metric (M = 80, SD = 10) across the span of 2nd–10th grades. This transformation allowed us to keep the integrity of any between-group or across-grade differences but put the variables on a common and, therefore, comparable metric. Because monotonic transformations such as this do not change the relative individual differences standings of the children either within time or across time, the developmental change patterns and associations are not artifactually altered. Our choice of 80 as the centering mean level and 10 as the standard deviation metric is arbitrary because there is no "standard" metric for all variables in the analyses.

Analyses were conducted within a structural equation modeling framework. This framework has the advantage of removing sources of measurement error and estimating indirect effects over time as well as modeling individual differences in intraindividual change (i.e., LGC analyses). Initially, we undertook a two-group longitudinal CFA to examine the measurement properties of the constructs of word recognition and reading comprehension as well as the stability of the constructs across grades. Next, we utilized LGC analyses to examine differences in the growth of word recognition and reading comprehension between the LI and TL groups (for a clear introduction to state-of-the-science CFA methods and their merits, see Brown, 2006).

All analyses were conducted with the LISREL 8.72 statistical package (Jöreskog & Sörbom, 2005) using maximum likelihood estimation. In these analyses, various models are tested for goodness of fit. The fit indices used to evaluate models were the following: the root-mean-square error of approximation (RMSEA), the non-normed fit index (NNFI), and the comparative fit index (CFI). For the RMSEA, values less than or equal to .08 are preferred. Values greater than .90 are generally considered to be acceptable for the NNFI and CFI. We report the chi-square fit statistic for all models. Although this statistic is highly sensitive to sample size, which precludes its use as an appropriate measure of goodness of fit in studies with sample sizes comparable with ours (see Jöreskog & Sörbom, 1993), we used the chi-square statistic to examine the adequacy of any nested comparisons regarding parameter estimates in the models. For these nested model chi-square difference tests, we used a p value of <.001 as the criteria for significance because of the extreme power present in our sample. We chose .001 because this is the p value that a sample size of 600 would give to detect a small effect (e.g., correlation of about .10 or a d of about 0.2), with a slight correction for multiple test error inflations. For the tests of measurement invariance, we used the change in CFI of >.01 rule as the criterion for failure of invariance (Cheung & Rensvold, 2002).

Results

CFAs

Multiple group CFAs were conducted to assess the extent to which the latent reading constructs were measured comparably across groups of children (LI and TL) and across grades (2nd, 4th, 8th, and 10th). Word recognition and reading comprehension were examined as separate latent constructs, and each construct comprised three indicators at each grade. In these CFA models, we included all assessed indicators of each construct, even those that were not assessed at all grades. We included the additional measures (DAB-2, QRI-2) as indicators of comprehension to get a complete picture of the comprehension construct over group and grade. For the invariance tests, we equated all corresponding loadings and intercepts of comparable measures (Brown, 2006). Table 1 displays the means and standard deviations for each of the indicators by grade and group after imputation but before conversion to the common metric we use in the LGC models (i.e., centered at $M = 80$, with $SD = 10$).

In the initial configural invariance model, each latent construct (word recognition or reading comprehension) was associated with the expected sets of indicators across groups and grades. The variances of the constructs were set to 1.0, and all intercorrelations among the constructs were freely estimated. Our configural invariant model fit the data well (see Table 2), indicating that the pattern of loadings for each latent construct was similar across groups and grades. Next, measurement invariance was evaluated by equating the loadings and intercepts in a series of sequential steps. First, we evaluated the equality of the factor loadings (i.e., the contributions of each indicator to the construct) across groups and grades. Second, the equality of indicator intercepts was examined across groups and grades. The former is a test of weak factorial invariance, and the latter is a test of strong factorial invariance. As shown in Table 2, each of the models had acceptable fit. These results indicate that the constructs included in the model (i.e., word recognition and reading comprehension) are invariant across grades and groups of children with LI or TL.

Next, we tested the homogeneity of the variances and covariances of the latent constructs in the LI and TL groups. If this omnibus test is nonsignificant, it indicates that the interrelations among the constructs do not differ across the two groups (and therefore a multiple-group comparison framework would not be needed). However, this test indicated there were significant differences in the variances and covariances between the two groups of children. Follow-up testing showed that differences in covariances were driving the significant omnibus test but that the variances did not differ across groups (see Table 2). Results of the test of homogeneity of the variances and covariances indicated that there were enough differences between the groups to proceed with a multiple-group LGC analysis instead of an analysis that collapses the two groups.

LGC Analysis

LGC analysis was undertaken to examine the differences in growth of word recognition and reading comprehension between children with LI and those with TL. The constructs identified in the previous analyses served as the basis for this analysis, with one exception. For the LGC analysis, the reading comprehension construct included only the WRMT-R Passage Comprehension subtest and the Comprehension standard score from the GORT-3. Comprehension scores from the DAB-2 and QRI-2 were not included because these measures were not administered across all four grades. That is, we took the average of the three common measures of word recognition

Table 1. Imputed means and standard deviations for the indicators of word recognition and reading comprehension.

	Children with typical language								Children with language impairment							
	Grade 2		Grade 4		Grade 8		Grade 10		Grade 2		Grade 4		Grade 8		Grade 10	
Construct indicator	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Word recognition																
WRMT–R Word ID	453.1	29.1	484.5	22.9	511.7	17.6	518.8	16.1	433.3	28.3	468.5	26.6	496.4	23.0	504.5	22.7
WRMT–R Word Attack	483.7	18.1	497.1	15.0	504.5	11.9	505.8	10.9	471.3	17.9	486.8	17.2	496.5	15.1	497.7	14.1
GORT–3 Accuracy	5.9	5.3	13.2	8.8	26.0	13.6	32.3	14.6	3.1	2.9	8.7	7.8	17.9	11.5	22.2	13.7
Reading comprehension																
WRMT–R Passage Comprehension	476.5	16.9	494.6	14.0	512.0	13.4	517.8	12.3	462.5	18.0	482.8	16.3	498.1	15.8	505.0	15.0
GORT–3 Comprehension	14.3	9.2	23.3	9.1	33.6	12.0	39.4	12.7	7.5	5.9	16.0	9.0	24.4	10.9	28.3	12.2
DAB–2 Comprehension	14.6	5.0	19.2	4.9					10.2	4.8	15.3	4.8				
QRI–2 Comprehension					9.1	4.2	12.0	3.5					5.9	3.7	8.4	4.2

Note. *W* scores are provided for the Woodcock Reading Mastery Tests–Revised (WRMT-R), and raw scores are listed for the other measures. *W* scores are on an interval scale that is centered at 500, which is the average performance level of beginning fifth graders. ID = Identification; GORT-3 = Gray Oral Reading Test–Third Edition; DAB-2 = Diagnostic Achievement Battery–2; QRI-2 = Qualitative Reading Inventory–Second Edition.

Table 2. Fit information for the steps to establish measurement invariance across grades and groups.

Model	χ^2	RMSEA	CI for RMSEA	NNFI	CFI
Configural invariance	757.16	.058	.052–.064	.989	.989
Weak invariance—across grades	858.92	.062	.056–.067	.988	.991
Weak invariance—across groups	943.80	.067	.061–.072	.986	.990
Strong invariance—across grades	877.33	.060	.054–.065	.991	.991
Strong invariance—across groups	1,032.99	.067	.062–.073	.985	.989
Homogeneity of variances/covariances	1,471.39	.078	.073–.082	.984	.981
Homogeneity of variances	999.96	.069	.063–.074	.989	.991

Note. RMSEA = root-mean-square error of approximation (preferred value is less than or equal to .08); CI = confidence interval; NNFI = nonnormed fit index (preferred value is greater than .90); CFI = comparative fit index (preferred value is greater than .90).

and the two common measures of reading comprehension as estimates of level and shape in the LGC models.

We specified a two-group multivariate level and shape model with the slopes of word recognition and reading comprehension fixed at 0 for 2nd grade and at 1 for 10th grade (Little, Bovaird, & Slegers, 2006). Here, the loadings (also termed *basis weights*) for 4th and 8th grades were freely estimated for the shape construct to show the unrestricted growth in reading over time. In these models, the level (or intercept) is interpreted as the initial status at 2nd grade. The shape factor represents the change in the reading achievement growth trajectory found across the grades. The level and shape model allows the shape of the trajectory to be optimally estimated, which could uncover nonlinearity in growth if present. All intercept and slope factors were allowed to covary within each group.

The initial model, which allowed unique basis weights (loadings) for the shape factor across groups, showed very good fit, $\chi^2(36, N = 604) = 107.20$; RMSEA = .082 (.065; .100); NNFI = .986; CFI = .991. We then tested whether the underlying trajectory of the shape factor was equivalent across the two groups by equating the estimated basis weights across the two groups. This model fit, $\Delta\chi^2(4, N = 604) = 17.20$, $p = .002$, failed to reach our predetermined level of significance ($p = .001$). Whereas it approached this level (because of our extreme power), the differences between groups in the shape of growth were quite small and functionally the same.[1] Inspection of the loadings in Figure 1 indicates a slight curvilinear pattern of growth, with more growth occurring between 2nd and 4th grades than between 4th and 10th grades. This pattern was the case for both word recognition and reading comprehension.

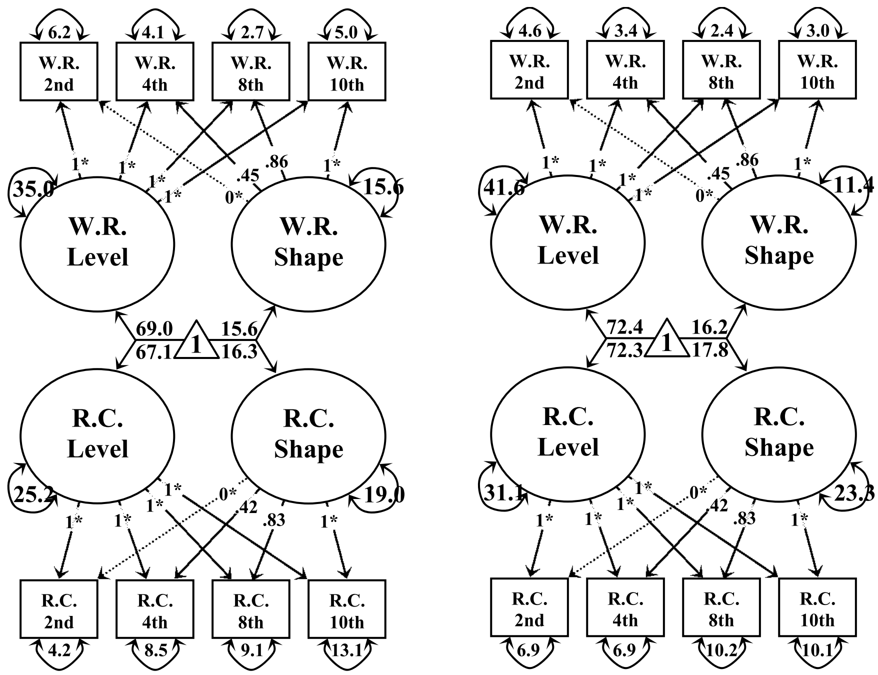
In the previous analyses, we used changes in basis weights to examine group differences. These differences were further evaluated in direct comparisons of the latent construct means. Specifically, we tested whether the initial levels and the magnitudes of the slopes of each construct were equal for children with LI and those with TL. An omnibus test revealed group differences, $\Delta\chi^2(4, N = 604) = 165.39$, $p < .0001$, so follow-up tests were conducted. In these tests, the individual means for the level or shape factor were compared across groups. Results indicate that the groups differed significantly in their intercept for both word recognition, $\Delta\chi^2(1, N = 604) = 61.77$, $p > .0001$, and reading comprehension, $\Delta\chi^2(1, N = 604) = 107.13$, $p > .0001$, with the LI group having significantly lower means (see Figure 1) than the TL group. The two groups did not differ, however, in the slope of their growth for either word recognition, $\Delta\chi^2(1, N = 604) = 2.74$, $p = .025$, or reading comprehension, $\Delta\chi^2(1, N = 604) = 8.80$, $p = .003$. This pattern of results indicates that the deficit in the LI group is present at second grade but that the changes thereafter are essentially parallel for the groups. This pattern of performance is further illustrated in Figure 2 for word recognition and in Figure 3 for reading comprehension.[2]

Finally, we entered the composite measure of nonverbal cognitive abilities to the LGC model as a covariate. Our results show that the model including this covariate continued to show acceptable fit, $\chi^2(48, N = 604) = 139.32$; RMSEA = .083 (.068; .98); NNFI = .984; CFI = .99. This model indicated that the composite measure of nonverbal cognitive abilities was moderately related to the intercept of reading achievement but essentially unrelated to

[1]The basis weights for each of the slope factors, which indicate the general shape or rate of growth, showed highly similar patterns. In the case of reading comprehension at Time 2, the basis weight increased to .452 of the total growth in the TL group and .407 in the LI group. At Time 3, the TL group increased to .850 of the total growth, whereas the LI group increased to .815. We found a similar pattern in word recognition, in which, at Time 2, the basis weight for the TL group increased to .477 and that for the LI group increased to .438, and, at Time 3, the basis weight for the TL group increased to .880 and that of the LI group increased to .855. Although we are aware of no formal effect size measure to summarize the magnitude of these differences, these basis weights are akin to regression slopes. Differences in regression slopes of the magnitude that we observed (e.g., .880 vs. .855) would be considered quite small.

[2]These figures are based on observed data rather than model-implied estimates. The slight differences between groups in the slopes from these raw estimates are not statistically significant.

Figure 1. Growth model including the constructs of word recognition (W.R.) and reading comprehension (R.C.) for children with language impairments (left) and typical language (right). Level = performance at second grade; Shape = the slope of the growth trajectory across grades. *Indicates that the parameter is fixed for identification and scaling purpose.



growth. For children with LI, the beta weights for the relationship between nonverbal cognitive abilities and the intercepts for word recognition and reading comprehension were .38 and .44, respectively. Similar beta weights for children with TL were .27 and .42. For children with LI, the beta weights for the relationship between nonverbal cognitive abilities and the slopes for word recognition and reading comprehension were .09 and .04, respectively. Similar values for children with TL were .06 and .03.

Discussion

In this study, we used LGC analysis to compare the pattern of reading growth from 2nd through 10th grades of children with LI to that of children with TL. Models

Figure 2. Growth in word recognition of children with language impairments (LI) and typical language (TL).

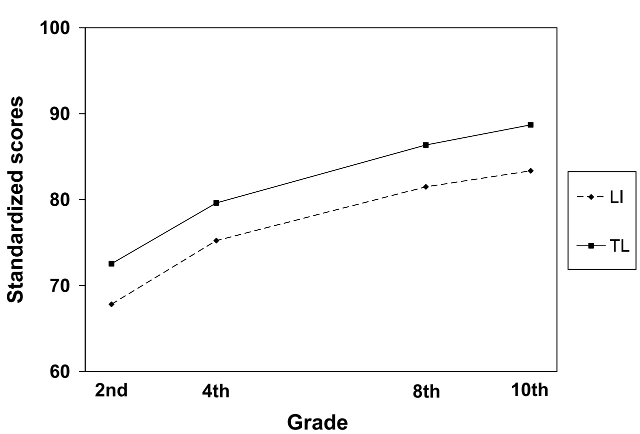
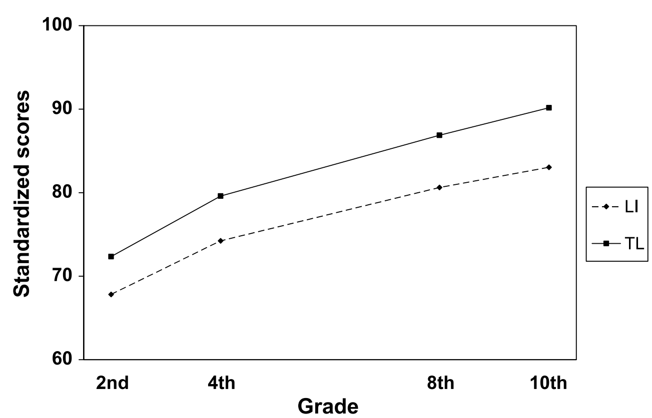


Figure 3. Growth in reading comprehension in children with language impairments (LI) and typical language (TL).



for both word recognition and reading comprehension showed that children with LI had a significantly lower intercept than did children with TL. Such a finding is evidence of early reading problems in children with LI, and it is consistent with our previous analyses involving this sample (Catts et al., 2002) and many other studies of children with LI (Bishop & Adams, 1990; Menyuk et al., 1991; Naucler & Magnusson, 1998; Roth et al., 2002; Silva et al., 1987; Tallal et al., 1989). Whereas LGC models indicated group differences in reading achievement intercepts, no significant group differences were observed in the shape of the growth trajectories. For word recognition and reading comprehension, both groups showed high initial acceleration followed by slower growth between 4th and 8th/10th grades. This growth pattern is consistent with that reported by Francis et al. (1996) in terms of its general shape. However, a direct comparison between our results and their findings concerning the grade at which growth deceleration occurred is not possible. Francis et al. conducted yearly assessments from 1st through 9th grades, whereas we had 2- or 4-year gaps in our measurement of reading achievement. Nevertheless, our finding of a slowing in the reading growth trajectories between 4th and 8th/10th grades is consistent with their report of a plateau in reading achievement growth near 6th grade.

Our results concerning group differences in intercept, but similarities in the slope of reading achievement growth trajectories, support a deficit model of reading achievement in children with LI. These children showed lower initial reading abilities and failed to catch up with their TL peers over the span of this investigation. In addition, group differences did not increase across grades, and thus, there was no evidence of a Matthew effect in reading achievement. Again, our findings are compatible with Francis et al. (1996), who reported a deficit model for poor readers. Whereas poor readers in their study were not identified on the basis of problems in language development, these children showed lower initial reading scores but growth trajectories similar to that of good readers.

Our findings of continued, long-term deficits in the reading achievement of children with LI are consistent with the few other studies that have examined reading achievement into the later grades in this population (Johnson et al., 1999; Snowling et al., 2000; Young et al., 2002). These results indicate that as a group, children with LI reach much lower levels of reading achievement in middle and high school grades than do children with TL. Our results from the 10th grade administration of the WRMT-R provide some indication of the extent of the reading achievement deficit in children with LI. The WRMT-R (Woodcock, 1987) provides nationally normed grade-equivalent scores in word reading and reading comprehension. Results from the 10th grade administration of this test show that children with LI performed at a grade-equivalent level of approximately 6th grade in reading comprehension and near 5th grade in word recognition. Children with TL were on grade level in reading comprehension but were below grade level in word recognition (7th–8th grades).[3]

Finally, when the composite measure of nonverbal cognitive abilities was added as a covariate into the model, it was found to be related to intercept but not shape of the reading achievement trajectories. These results indicate that children with low nonverbal cognitive abilities (in both groups) generally had lower initial reading achievement than those with normal nonverbal cognitive abilities but showed no differences in the growth of reading achievement over the school grades. Nonverbal cognitive abilities are not generally thought to impact reading achievement as much as verbal abilities (Stanovich, 1991). However, it is possible that that the analytical reasoning and visual–spatial skills assessed by our nonverbal cognitive measures could influence both word reading and reading comprehension. Our findings concerning nonverbal cognitive abilities are in line with the results of other studies that have reported that children with LI and low nonverbal IQ have poorer reading outcomes than those with normal nonverbal IQ (Bishop & Adams, 1990; Snowling et al., 2000). On the other hand, our results are contrary to those of Snowling and colleagues (2000), who found a tendency for children with lower nonverbal cognitive abilities to show slower growth in word reading than those with normal abilities. However, conclusions from the latter study are limited by a less than optimal design for assessing growth.

A limitation of our study was that we did not assess reading achievement prior to 2nd grade. Thus, we are unable to evaluate the reading achievement growth pattern of children with LI from kindergarten/1st grade to 2nd grade. It is possible that the pattern of growth during that period might have been different from what we observed from 2nd to 10th grades. Recall that Bast and Reitsma (1998) reported a cumulative growth pattern or Matthew effect from kindergarten through 2nd grade

[3]It is unclear why our TL sample achieved lower than expected levels in word recognition but appropriate levels in reading comprehension in 10th grade. This may have resulted from a curriculum that was in place at the time of our study in which less attention was focused on phonics. The fact that the TL group's average performance was lower on the Word Attack subtest (an index of phonics knowledge) than the Word Identification subtest seems to support this conclusion. However, the impact of the curriculum would have to have been cumulative because the TL group did achieve grade-level scores in word reading in the early school grades (2nd, 4th) before falling behind national norms in the later grades. An additional possibility is that the norms for the word reading and reading comprehension subtests were out of sync. A normative update for the WRMT-R (Woodcock, 1998) provides some support for this possibility. This update converts the TL group's mean *W* score for the Passage Comprehension subtest (10th grade) to a grade level of 8.3, a value more comparable with that found for the Word Identification conversion (7.6).

in children from the general population. The first author and colleagues (Puranik, Petscher, Al Otaiba, & Catts, in press) have reported some evidence of a cumulative growth pattern during 1st grade in another sample of children with LI. In this study, we selected children who were identified in a statewide database as having LI in 1st, 2nd, and/or 3rd grades. These children were further subdivided into those whose problems had resolved (no longer designated) by 2nd or 3rd grades and those whose problems were persistent across all three grades. Using similar procedures, we also selected children with a persistent or resolved speech impairment. We then conducted growth curve analyses on children's oral reading fluency, which had been assessed four times a year from the beginning of 1st grade through the end of 3rd grade. Our results show a cumulative pattern of growth for groups across 1st grade. Specifically, those children with a persistent impairment, especially persistent LI, showed an initial deficit and slower growth than other groups. In 2nd grade and throughout 3rd grade, however, groups continued to differ in reading fluency but showed a similar rate of growth (i.e, deficit pattern). Because this study utilized a secondary data source, it is not clear how comparable the children with LI in that study are to those in the current investigation. However, it does suggest the need for further investigation of reading growth among children with LI in the early school grades.

Implications

The observation that our data best fit a deficit model for reading achievement has important implications for early identification. That is, given that children who start out as poor readers remain poor readers across the school grades, it is critical to identify these children early and provide appropriate intervention to reduce the long-term consequences of reading problems. The results of this study and several others (Johnson et al., 1999; Snowling et al., 2000; Young et al., 2002) demonstrate that the presence of LI in kindergarten is an important indicator of a reading disability in the later school grades. Therefore, screening and referral systems need to be in place to identify children with LI at the beginning of kindergarten or earlier. Whereas some schools have implemented procedures to identify children with LI, much of the current attention in early identification in the schools is focused on assessment in areas such as phonological awareness and letter knowledge (Foorman et al., 1998; Good & Kaminski, 2002). Results from our previous analyses (Catts et al., 2002; Catts, Fey, Zhang, & Tomblin, 2001) confirm the importance of such assessments. However, the current results also suggest the need to go beyond these traditional early literacy variables and consider other aspects of language development. Screenings need to consider early language skills in vocabulary, grammar, and/or narration to identify children who are at risk for a reading disability on the basis of poor language skills. In some cases, these children's language skills will be low enough to identify them as having LI. However, recent investigations indicate that some children who are at risk for a reading disability on the basis of language deficits may show more moderate deficits and do not qualify for an LI designation in kindergarten. These studies have investigated children referred to as "poor comprehenders" (Catts, Adlof, & Ellis-Weismer, 2006; Nation, Clarke, Marshall, & Durand, 2004; Oakhill & Yuill, 1996). These are children who have significant problems in reading comprehension despite adequate or better word recognition skills. They may also be children whose reading problems are late to emerge (Leach, Scarborough, & Rescorla, 2003). Results show that most poor comprehenders have early language deficits. However, only a small percentage of these children have severe enough deficits to have been identified as having LI. Thus, screening assessments need to be able to identify children with LI as well as those with more moderate deficits.

Acknowledgments

This research was supported by National Institute on Deafness and Other Communication Disorders Grant 1-P50-DC02726-04 and also by National Institutes of Health (NIH) Training Grant T32 DC000052 that supported the second author. This research was also supported in part by grants from the NIH to the University of Kansas through the Mental Retardation and Developmental Disabilities Research Center (5 P30 HD002528) and the Center for Biobehavioral Neurosciences in Communication Disorders (5 P30 DC005803). The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The authors would like to thank Xuyang Zhang, Marc Fey, Susan Ellis Weismer, Paula Buckwalter, Marlea O'Brien, Connie Ferguson, Jodi Schwartz, Amy Schminke, Marcia St. Clair, Wendy Fick, and Juanita Limas for their valuable contributions to this investigation.

References

Baker, L. A., Decker, S. N., & DeFries, J. C. (1984). A longitudinal study of cognitive abilities in reading-disabled and control children. *Journal of Child Psychology and Psychiatry, 25,* 111–117.

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology, 34,* 1373–1399.

Beitchman, J. H., Nair, R., Clegg, M., & Patel, P. G. (1986). Prevalence of speech and language disorders in 5-year-old kindergarten children in the Ottawa–Carleton region. *Journal of Speech and Hearing Disorders, 51,* 98–110.

Bishop, D. V.M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry, 31,* 1027–1050.

Botting, N., Simkin, Z., & Conti-Ramsden, G. (2006). Associated reading skills in children with a history of specific language impairment (SLI). *Journal of Reading and Writing, 19,* 77–98.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research, 36,* 948–958.

Catts, H. W., Adlof, S., & Ellis-Weismer, S. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49,* 278–293.

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45,* 1142–1157.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3,* 331–361.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32,* 38–50.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255.

Culatta, B., Page, J., & Ellis, J. (1983). Story retelling as a communicative performance screening tool. *Language, Speech, and Hearing Services in Schools, 14,* 66–74.

Foorman, B. F., Fletcher, J. M., Francis, D. J., Carlson, C. D., Chen, D., Mouzaki, A., et al. (1998). *Texas Primary Reading Inventory* (1998 ed.). Houston, TX: Center for Academic and Reading Skills and University of Houston.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88,* 3–17.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Retrieved October 16, 2006, from http://dibels.oregon.edu/.

Graham, J. W., & Schafer, J. L. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Hayes, D. P., & Grether, J. (1983). The school year and vacations: When do students learn? *Cornell Journal of Social Relations, 17,* 56–71.

Johnson, C. J., Beitchman, J. H., Young, A. R., Escobar, M., Atkinson, L., Wilson, B., et al. (1999). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research, 42,* 744–760.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: SSI.

Jöreskog, K. G., & Sörbom, D. (2005). *LISREL*. Lincolnwood, IL: SSI.

Juel, C. (1988). Learning to read and write: A longitudinal study of fifty-four children from first through fourth grade. *Journal of Educational Psychology, 80,* 437–447.

Kaufman, A. S. (1979). *Intelligence testing with the WISC-R*. New York: Wiley.

Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology, 95,* 211–224.

Leppänen, U., Niemi, P., Aunola, K., & Nurmi, J. E. (2004). Development of reading skills among preschool and primary school pupils. *Reading Research Quarterly, 39,* 72–93.

Leslie, L., & Caldwell, J. (1995). *Qualitative Reading Inventory–2*. New York: Harper Collins.

Little, T. D., Bovaird, J. A., & Slegers, D. W. (2006). Methods for the analysis of change. In D. K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 181–211). Mahwah, NJ: Erlbaum.

McKinney, J. D., & Feagans, L. (1984). Academic and behavioral characteristics: Longitudinal studies of learning disabled children and average achievers. *Learning Disability Quarterly, 7,* 251–265.

Menyuk, P., Chesnick, M., Liebergott, J. W., Korngold, B., D'Agostino, R., & Belanger, A. (1991). Predicting reading problems in at-risk children. *Journal of Speech and Hearing Research, 34,* 893–903.

Nation, K., Clarke, P., Marhsall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairments? *Journal of Speech, Language, and Hearing Research, 47,* 199–211.

Naucler, K., & Magnusson, E. (1998). Reading and writing development: Report from an ongoing longitudinal study of language-disordered and normal groups from pre-school to adolescence. *Folia Phoniatrica et Logopaedica, 50,* 271–282.

Newcomer, P. (1990). *Diagnostic Achievement Battery–2*. Austin, TX: Pro-Ed.

Newcomer, P., & Hammill, D. (1988). *Test of Language Development–2: Primary*. Austin, TX: Pro-Ed.

NICHD Early Child Care Research Network. (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology, 41,* 428–442.

Oakhill, J. V., & Yuill, N. (1996). Higher order factors in comprehension disability: Processes and remediation. In C. Cornoldi & J. V. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 69–92). Mahwah, NJ: Erlbaum.

Puranik, C., Petscher, Y., Al Otaiba, S., & Catts, H. (in press). How do speech or language impairments effect reading fluency development? A growth curve analysis. *Journal of Learning Disabilities*.

Roth, F. P., Cooper, D. H., & Speece, D. L. (2002). A longitudinal analysis of the connection between oral language and early reading. *Journal of Educational Research, 95,* 259–272.

Scarborough, H. S., & Parker, J. D. (2003). Matthew effects in children with learning disabilities: Development of reading, IQ, and psychosocial problems from Grade 2 to Grade 8. *Annals of Dyslexia, 53,* 47–71.

Shaywitz, B. A., Holford, T. R., Holahan, J. M., Fletcher, J. M., Stuebing, K. K., Francis, D. J., et al. (1995). A Matthew effect for IQ but not for reading: Results from a longitudinal study. *Reading Research Quarterly, 30,* 894–906.

Silva, P. A., McGee, R., & Williams, S. R. (1987). Developmental language delay from 3 to 7 years and its significance for low intelligence and reading difficulties at age seven. *Developmental Medicine and Child Neurology, 25,* 783–793.

Snowling, M., Bishop, D. V. M., & Stothard, S. E. (2000). Is preschool language impairment a risk factor for dyslexia in adolescence? *Journal of Child Psychology and Psychiatry, 41,* 587–600.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360–406.

Stanovich, K. E. (1991). Discrepancy definitions of reading disability: Has intelligence led us astray? *Reading Research Quarterly, 26,* 7–29.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38,* 934–937.

Tallal, P., Curtis, S., & Kaplan, R. (1989). *The San Diego Longitudinal Study: Evaluating the outcomes of preschool impairment in language development: Final report*. Washington, DC: National Institute of Neurological and Communication Disorders and Stroke.

Tomblin, J. B. (1995). *Midwest collaboration on specific language impairment*. Washington, DC: National Institute on Deafness and Other Communication Disorders.

Tomblin, J. B., Records, N., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). The prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40,* 1245–1260.

Tomblin, J. B., Records, N., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research, 39,* 1284–1294.

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence–Revised*. New York: Psychological Corporation.

Wiederholt, J., & Bryant, B. (1992). *Gray Oral Reading Test* (3rd ed.). Austin, TX: Pro-Ed.

Williamson, G. L., Appelbaum, M., & Epanchin, A. (1991). Longitudinal analyses of academic achievement. *Journal of Educational Measurement, 28,* 61–76.

Woodcock, R. (1987). *Woodcock Reading Mastery Tests–Revised*. Circle Pines, MN: American Guidance Service.

Woodcock, R. (1998). *Woodcock Reading Mastery Tests–Revised/Normative Update*. Circle Pines, MN: American Guidance Service.

Young, A. R., Beitchman, J. H., Johnson, C., Douglas, L., Atkinson, L., Escobar, M., et al. (2002). Young adult outcomes in a longitudinal sample of early identified language impaired children and control children. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43,* 635–645.

Received November 21, 2007

Accepted April 27, 2008

DOI: 10.1044/1092-4388(2008/07-0259)

Contact author: Hugh W. Catts, Department of Speech, Language, Hearing, 1000 Sunnyside Avenue, University of Kansas, Lawrence, KS 66045. E-mail: catts@ku.edu.