JSLHR

Research Article

Predictors of Language Gains Among School-Age Children With Language Impairment in the Public Schools

Laura M. Justice,[a] Hui Jiang,[a] Jessica A. Logan,[a] and Mary Beth Schmitt[b]

Purpose: This study aimed to identify child-level characteristics that predict gains in language skills for children with language impairment who were receiving therapy within the public schools. The therapy provided represented business-as-usual speech/language treatment provided by speech-language pathologists in the public schools.

Method: The sample included 272 kindergartners and first-graders with language impairment who participated in a larger study titled "Speech-Therapy Experiences in the Public Schools." Multilevel regression analyses were applied to examine the extent to which select child-level characteristics, including age, nonverbal cognition, memory, phonological awareness, vocabulary, behavior problems, and self-regulation, predicted children's language gains over an academic year. Pratt indices were computed to establish the relative importance of the predictors of interest.

Results: Phonological awareness and vocabulary skill related to greater gains in language skills, and together they accounted for nearly 70% of the explained variance, or 10% of total variance at child level. Externalizing behavior, nonverbal cognition, and age were also potentially important predictors of language gains.

Conclusions: This study significantly advances our understanding of the characteristics of children that may contribute to their language gains while receiving therapy in the public schools. Researchers can explore how these characteristics may serve to moderate treatment outcomes, whereas clinicians can assess how these characteristics may factor into understanding treatment responses.

Language impairment (LI), whether affecting a child as a primary or secondary disability, is one of the most common developmental disabilities for which children receive specialized services in the early primary grades (U.S. Department of Education, 2015). Services for LI are provided within the public schools via federal funding sources because of the well-demonstrated adverse effects of LI on a variety of key educational outcomes, including children's social relations with peers (Laws, Bates, Feuerstein, Mason-Apps, & White, 2012) and their reading and mathematics skills (Morgan, Farkas, & Wu, 2011; Skibbe et al., 2008). In general, the primary goal of language therapy provided to children with LI in the primary grades is to accelerate their language growth and to strengthen those skills most critical to academic success.

Children with LI are often described homogeneously, although they are in reality a heterogeneous population (Conti-Ramsden & Botting, 1999; Tambyraja, Schmitt, Farquharson, & Justice, 2015). Some efforts to classify these youngsters have focused on identifying the extent to which specific language dimensions, such as grammar and phonology, are affected (e.g., Conti-Ramsden & Botting, 1999), whereas others have shown heterogeneity to be largely a function of the severity of the disability (see Tambyraja et al., 2015) or the co-occurrence of nonverbal deficits (e.g., Catts, Fey, Tomblin, & Zhang, 2002). Understanding heterogeneity among this population is often framed theoretically, but it also likely has consequences to the treatment of children with LI. That is, treatment studies show that children's gains in language skills within the context of treatment can be highly variable, with some children gaining more in language skill than others (Gillam et al., 2008; Petersen, Gillam, Spencer, & Gillam, 2010; Washington, Thomas-Stonell, McLeod, & Warr-Leeper, 2015). For instance, an experimental investigation of an

[a]The Ohio State University, Columbus
[b]Texas Tech University, Lubbock

Correspondence to Laura M. Justice: justice.57@osu.edu

Editor: Sean Redmond
Associate Editor: Patricia Eadie

Received January 21, 2016
Revision received June 30, 2016
Accepted November 11, 2016
https://doi.org/10.1044/2016_JSLHR-L-16-0026

Disclosure: The authors have declared that no competing interests existed at the time of publication.

intensive language intervention for children with specific LI (representing a subpopulation of children with language disorders absent of intellectual impairment) showed large standard deviations on standardized measures of language skill at posttest and several months postintervention (e.g., $M = 28$, $SD = 12$ on a comprehension task; Gillam et al., 2008).

To date, we have limited understanding of heterogeneity among children with LI in their gains in language skill over time, although it is likely that certain child-level characteristics have important prognostic value in explaining variability among children in their language gains over time (Bishop, Snowling, Thompson, Greenhalgh, & CATALISE Consortium, 2016). For instance, a recent study of preschoolers with speech-language impairments receiving community-based treatment found that children's social competence was a significant and negative predictor of change scores on a language assessment; children with poorer social skills showed less change in language skills during interactions than children with better social skills (Washington et al., 2015). Such work suggests that social competence may be an important prognostic indicator of children's language development over time, especially for children receiving speech-language services.

The purpose of the present study was to identify the extent to which certain child-level characteristics can explain the variance in their language development over an academic year during which the children were receiving language services in the public schools. Although we conceptualize these language services as representing an intervention being received by the children, it is important to point out that the intervention was not scripted or prescribed in any way. Rather, it represented speech-language pathologists' (SLPs) business-as-usual (BAU) practices when providing services to children with LI, and there was considerable variability among SLPs in various parameters related to the services they provided (Farquharson, Tambyraja, Logan, Justice, & Schmitt, 2015). We theorized that child-level characteristics that generally predispose children to improved academic achievement, including their cognition, short-term memory, positive behavior, and self-regulation, might also be associated with variability in the language gains of children with LI when receiving language-related services in the schools.

Often, research on the treatment of childhood language disorders is focused on identifying effective avenues for bringing about change in specified language outcomes, such as vocabulary (Throneburg, Calvert, Sturm, Paramboukas, & Paul, 2000), morphosyntax (Camarata, Nelson, & Camarata, 1994), narrative skills (Petersen et al., 2010), and phonological awareness (Gillon, 2002). A complementary area of treatment research involves examining whether treatment effects are conditional on (i.e., moderated by) specific characteristics of children. Put differently, language intervention may be particularly beneficial to some children as a function of certain characteristics, but less so for others. If this were the case, treatment effects in a larger-scale trial would likely be difficult if not impossible to detect without consideration of these pretreatment characteristics as moderators of outcomes (i.e., via analysis of specific child × treatment interactions). Child × treatment interactions capture the premise that individual differences should be taken into account during evaluations of treatments (Dance & Neufeld, 1988; Snow, 1991). For instance, highly intensive language treatments may be effective only for children with high levels of self-regulation (as these children are better able to manage their own behavior); if this were the case, children's self-regulation would be considered a moderator of treatment and would need to be modeled in relation to the treatment to discern such effects.

The potential for a given child-level characteristic to moderate treatment outcomes is substantially increased when children show a great deal of variability on that skill, referred to as an individual difference. An example of an individual difference long viewed to be of interest in the study of children with LI is nonverbal cognition, with children grouped into those with normal levels of nonverbal cognition, called a primary or specific language disorder, and those with below-normal levels, called a nonspecific language disorder (Catts et al., 2002). Children in the former group typically exhibit a significant discrepancy between language and cognition, whereas children in the latter group have no such discrepancy, with low levels of language skill commensurate with low levels of cognition. It is unclear whether clinicians in the public schools today utilize cognitive referencing in their clinical decisions regarding eligibility and service-delivery approach; at least historically, however, there were suggestions that children with a primary condition would benefit more from language intervention than those with a nonspecific condition and thus should be considered prime candidates for intervention (see Cole, Mills, & Kelley, 1994). In a recent article, Bishop argued that nonverbal cognition places no limits on language development and, accordingly, has little utility in understanding language gains among children with LI or their treatment prognosis (Bishop, 2014). In the present study, we empirically examined whether the nonverbal cognition of children with LI receiving services within the public schools is a significant predictor of their gains over the academic year. The result of this examination may be useful for determining whether nonverbal cognition should be included as a moderator in treatment research, and whether nonverbal cognition serves as an important prognostic indicator of language gains among this population.

One of the challenges in studying child × treatment interactions within the context of a treatment study is that most studies are powered to identify main effects (i.e., the effect of the treatment on key outcomes), but not treatment moderators. Simply put, statistical power refers to the sensitivity of an experiment to accurately detect that the tested hypothesis is true and that the null should be rejected. In treatment research, the tested hypothesis is typically whether a given treatment has positive effects on children's language skills relative to the counterfactual, and a study is adequately powered when those effects can be identified at the expected level, even if they are quite small (e.g., effect size of .1 or .2).

Adequate statistical power to correctly interpret an experiment's results is based on several different factors, one of which is the number of participants to include. Often, when one is designing a treatment study, a priori decisions are made regarding statistical power, such as establishing the desired number of participants to be included in the study to detect whether the given treatment has positive effects. This number is usually determined on the basis of one's interest to correctly interpret the main effect of the experiment, and it is not used to test for potential moderators (which requires greater power than the test for the main effect). Thus, when an experiment is concluded, and one wants to determine whether the intervention was more or less effective given a particular pretreatment characteristic, such as children's nonverbal cognition, the sample size is usually too small to engage in such analyses appropriately and to detect the moderator effects of interest (e.g., Kouri, 2005). Nonetheless, the role of child-level pretreatment characteristics and other moderators can be paramount factors in understanding for whom and under what conditions a given treatment is effective.

An alternative to testing for moderation within a treatment study is to explore whether certain child-level characteristics are predictive of their gains in language skill while receiving language services in the schools. This is essentially correlational work, in which one tries to link certain child-level characteristics with the magnitude of language gain seen over time. Olswang and Bain used this approach to examine the extent to which certain characteristics of 2-year-old children were associated with the amount of change in language skill they exhibited over a 9-week period of intervention (Olswang & Bain, 1996). The strongest child-level predictor of change was performance on a dynamic assessment task (r = .73) designed to determine how much support was needed from an adult to elicit a two-word utterance. The study results suggested that toddlers who were poised to produce two-word utterances showed greater improvement in their language skills during treatment than those who were not. Such work is informative for pinpointing the moderators that should be assessed, of the many possible candidates available, in future treatment studies that are focused on late-talking toddlers.

The present study was designed to describe the variability in language gains among kindergarten and first-grade children and to identify child-level characteristics that help to understand this variability. The children were being provided language therapy within the public schools on the basis of the prevailing, BAU practices of their SLPs. Thus, this study looked at the variability in language gains over time for children receiving BAU practices, and therefore it represents an exploratory investigation rather than an experimental intervention. The child-level characteristics that we investigated included age, four cognitive skills (i.e., nonverbal cognition, short-term memory, phonological awareness, vocabulary), and three noncognitive skills (i.e., externalizing and internalizing behavior problems, self-regulation).

The rationale for selecting these eight variables was largely based on two factors. First, each variable represents a documented individual difference among children with LI (e.g., Archibald & Gathercole, 2006; Briscoe, Bishop, & Frazier Norbury, 2001) as well as children developing typically (Metsala, 1999; Ponitz, McClelland, Matthews, & Morrison, 2009). Given that children normally show variability on these indices, it stands to reason that this variability may help us to understand individual differences in language gains over time.

Second, for many of these variables, the extant empirical and theoretical literature has suggested that these may relate to gains during language therapy among children with LI; specifically, each characteristic may have unique, explanatory value for explaining variance in children's language gains during an academic year. First, age was included because evidence suggests that language gains during an academic year decrease as children get older. Schmitt et al. recently reported that 5-year-old children gain an average of 0.74 *SD* in language skill over an academic year, whereas 7-year-old children gain about 0.5 *SD* (Schmitt, Logan, Tambyraja, Farquharson, & Justice, 2017). We therefore anticipated that age would be negatively associated with language gains for children with LI. Second, nonverbal cognition was included because of the long-standing perception that it has prognostic value for understanding language gains among children with LI, recognizing that experts have argued that this perception should be discarded (see Bishop, 2014).

Third, short-term memory was included because of theoretical and empirical evidence concerning its role in language development, especially word learning (Côté, Rouleau, & Macoir, 2014; Majerus & Boukebza, 2013). We anticipated that children with higher levels of vocabulary skill would show greater language gains over time, accordingly. Fourth, phonological awareness was included because it plays a strong predictive role in understanding development of reading comprehension among children with LI during the primary grades (Catts et al., 2002). Given the strong relations between reading comprehension and language skills, we anticipated that children with higher levels of phonological awareness would show greater gains in language skill over time. Fifth, children's vocabulary skills were also included as a predictor variable, because studies have shown that children with high levels of vocabulary skill make greater language gains over time, representing a Matthew effect (Penno, Wilkinson, & Moore, 2002).

In addition, we included a robust set of noncognitive skills among our predictors, including externalizing and internalizing behavior problems and self-regulation. Some work has shown that noncognitive skills are negative prognostic indicators of language gains over time among children with LI (e.g., Washington et al., 2015). Researchers argue that such noncognitive factors are associated with children's language gains over time because they enable children to be attentive during learning situations (Schmitt, Pentimonti, & Justice, 2012).

To summarize, the aims of this study were twofold: (a) to examine gains in language skills over an academic year for children with LI, all of whom were receiving language intervention in the public schools, and (b) to identify the extent to which select characteristics of children were significant predictors of gains, and to establish the relative importance of the predictors of interest.

Method

Participants

The current study involved secondary analysis of data representing 272 children who participated in the study titled "Speech-Therapy Experiences in the Public Schools (STEPS)." STEPS featured a multicohort, prospective research design that involved ascertaining a clinically identified sample of children with LI from among SLP caseloads in the fall of the academic year; children were then followed intensively for 9 months to document (a) changes in skills of interest (e.g., language, reading) and (b) therapy experiences (e.g., treatment intensity, service-delivery model), as well as the relations between the two. The total sample involved 294 kindergarten and first-grade children nested within 75 SLP caseloads. The participant selection criteria and demographic information have been detailed in prior studies involving analysis of STEPS data (e.g., Farquharson et al., 2015; Tambyraja et al., 2015; Tambyraja, Schmitt, Justice, Logan, & Schwarz, 2014). As an overview, participant selection occurred in a two-step process. First, SLPs who served 5- to 7-year-old children with LI in the public schools were asked for their consent. Second, three to five children from each SLP's caseload who met the following initial criteria were asked for their consent for participation: (a) had provision of caregiver consent, (b) had an established diagnosis of LI and were receiving speech/language services, and (c) were proficient in English. For the present purposes, we applied a cutoff point of 85 (−1 *SD* of the mean) to the fall language assessment administered to each child by the project staff, eliminating data for 22 children in the STEPS study. By only including children who had a score of 85 or below, we ensured that the children in the present study represented children who had been identified by both the researcher and through a clinical diagnosis as having LI.

With respect to participant characteristics, the 75 SLPs were primarily women (99%), and all had their certificate of clinical competence from the American Speech-Language-Hearing Association as well as state licensure. On average, the SLPs had practiced for 16 years (*SD* = 11 years). For the 272 children, 88 (32%) were girls, and 184 (68%) were boys. The majority were Caucasian (54%, *n* = 147); 11% were African American (*n* = 29); 4% were Hispanic (*n* = 11); 7% were other ethnicities (*n* = 18); and the rest were unreported (*n* = 67). These children were, on average, 76 months old at the beginning of the academic year (*SD* = 8.5, range 59 to 96 months). As an indicator of socioeconomic status, 208 of the children's families reported their annual household income at intake (24% unreported): 32% of families earned less than $40,000 (*n* = 88), 27% earned between $41,000 and $80,000 (*n* = 74), and 16.9% earned more than $80,000 (*n* = 46). Table 1 provides these and other demographic details for the sample of children.

General Procedures

The participation of SLPs and children spanned one academic year of schooling. Prior to the start of the academic year, SLPs attended a 90-min orientation session in which they received instructions and all necessary materials for tracking the therapy experiences of children enrolled in the study. In particular, SLPs received materials to keep weekly written logs of treatment provision for each child enrolled in the study, and for collecting periodic videotapes of therapy sessions over the academic year; in total, SLPs would collect and submit videotapes for up to five therapy sessions per child in the study. In addition, the SLPs completed questionnaires about children's skills at various time points. The children completed a battery of assessments in the fall and spring of the academic year conducted by project staff during two 45-min sessions per time point. The battery included measures of nonverbal cognition, language, reading, and memory.

It is important to reiterate that STEPS examined children's gains in language skill while receiving BAU services within the public schools. Thus, there was considerable naturalistic variation among children in the nature of the services provided. The intent of STEPS was to explore this naturalistic variation and assess how this variability may relate to children's gains over time. In addition, the STEPS study provided the opportunity to examine predictors of language gains over time among this large, clinically relevant sample, as presented herein.

Measures

Measures collected in STEPS that are relevant to the present study included measures of children's language skills and pretreatment characteristics of nonverbal cognition, memory, phonological awareness, vocabulary, behavior problems, and self-regulation. In addition, measures related to the intensity of treatment received by each child were collected. Tables 2 and 3, respectively, provide descriptive data for this sample on their pretreatment characteristics and treatment intensity.

Children's Language Skills

Four subtests of the Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003) were administered to measure children's language skills at two time points (fall and spring): Concepts and Following Directions, Word Structure, Formulated Sentences, and Recalling Sentences. The four subtests are used to derive a core language composite score (raw and standard score). The subtests have adequate psychometric characteristics, including test–retest reliability and internal consistency (Semel et al., 2003). The CELF-4 fall and spring scores served as the primary indexes of treatment response in the present study; for our purposes, a simple

Table 1. Demographic characteristics of the sampled children.

Characteristic	All participants	Subsample[a]	
		Typical therapy	Shortened therapy
N	272	247	22
Mean age (*SD*) in months (baseline)	76.2 (8.5)	76.1 (8.5)	75.7 (8.0)
% female participants	32.4	34.0	18.2
Ethnicity			
% White/Caucasian	54.0	55.1	45.5
% African American	10.7	10.1	13.6
% Hispanic	4.0	3.6	9.1
% other	6.6	6.5	9.0
% unreported	24.6	24.7	22.7
Family income (categorical)			
% ≤ $40K	32.4	32.0	36.4
% $41K–$80K	27.2	27.9	22.7
% > $80K	16.9	17.0	13.6
% unreported	23.5	23.1	27.3
Mother's highest level of education			
% high school or lower	27.9	29.1	18.2
% some college, Associate of Arts/Associate of Science	27.6	26.3	40.9
% bachelor's degree	18.8	19.0	18.2
% master's or higher	5.9	6.1	0.0
% unreported	19.5	19.0	22.7

[a]Typical therapy schedule subsample consists of children who received between 15 and 80 therapy sessions over a span of 30 to 40 weeks (0.5–2.4 sessions per week). Shortened therapy schedule subsample consists of children who received fewer than 30 weeks of therapy. There are also three children who were exposed to extremely intensive treatment schedules (more than 100 sessions, 3–4 sessions per week), but the demographic statistics of this subgroup are not reported here due to the limited sample size.

gain score (change in the composite raw score from fall to spring) was calculated to represent improvements over time on this measure. Standard scores were not used in calculating the index of gain, because these were adjusted for age at each time point in their calculation.

Nonverbal Cognition

Children's nonverbal cognition was measured by the Matrices subtest of the Kaufman Brief Intelligence Test–Second Edition (KBIT-2; Kaufman & Kaufman, 2004). The subtest items use visual depictions of stimuli to assess a child's understanding of the relationships among stimuli. This measure was selected because it transcends the age range of the participant sample and can be administered briefly (about 5 to 10 mins, on average, per child). The Matrices subtest has adequate psychometric characteristics (e.g., test–retest reliability, internal consistency) as indicated by the administration manual. The subtest has a standardized mean of 100 and *SD* of 15, which allows a child's cognition to be compared with a normative sample. On average, the sampled children scored at the 26th percentile, with M = 88.4 and SD = 12.0.

Memory

Children's working memory was examined using the Number Repetition subtest of the CELF-4. In this subtest, children are asked to repeat series of number strings both forward and backward; strings repeated correctly are scored as correct, and these are summed to arrive at an overall subtest score. At baseline, the sampled children received a mean raw score of 5.8 (SD = 2.7).

Expressive Vocabulary

The Picture Vocabulary subtest of the Woodcock-Johnson III Test of Cognitive Abilities (WJ-III; Woodcock, McGrew, & Mather, 2001) was administered to provide an index of vocabulary skill in the fall of the year. The subtest requires the child to produce the name of pictured objects. WJ-III raw scores are transformed to *W* scores, which are derived from item-response-theory analyses. The mean *W* score for the present sample was 469.4 (SD = 11.4). The subtest manual indicates good reliability for the age of our sample (.88 for children ages 4 to 7 years). WJ-III raw scores were used for the current study.

Phonological Awareness

Children's phonological awareness was measured using the Catts Deletion Task, which is provided in the appendix of a published research report (Catts, Fey, Zhang, & Tomblin, 2001). Administered in the fall of the year, children responded to a series of increasingly complex segmentation tasks. Raw scores represent children's performance, with a possible score range of 0 to 21. The present sample of children received a mean raw score of 5.4 (SD = 7.1).

Problem Behavior

Two subscales of the Social Skills Rating System (SSRS; Gresham & Elliott, 1990), SSRS–Externalizing

Table 2. Descriptive statistics for children's pretreatment characteristics.

Characteristic	All participants	Subsample[a]	
		Typical therapy	Shortened therapy
N	272	247	22
CELF-4 Core Language			
CFD raw score	17.5 (10.6)	17.3 (10.7)	20.36 (10.2)
Word Structure raw score	13.5 (6.2)	13.2 (6.3)	16.6 (5.7)
Recalling Sentences raw score	21.6 (13.1)	21.4 (13.2)	25.8 (11.3)
Formulated Sentences raw score	8.6 (8.8)	8.3 (8.7)	12.9 (9.3)
Total Core Language raw score	61.2 (34.0)	60.3 (34.2)	75.7 (30.2)
Total Core Language standard score	68.8 (17.3)	68.1 (17.4)	79.7 (12.3)
Kaufman Brief Intelligence Test (KBIT)			
Raw score	15.4 (4.3)	15.3 (4.3)	16.7 (4.5)
Standard score	88.4 (12.0)	88.1 (12.0)	92.0 (12.1)
Percentile rank	26.5 (21.9)	26.1 (21.6)	32.3 (25.5)
CELF-4 Memory raw score	5.8 (2.7)	5.8 (2.7)	7.0 (2.1)
WJ-III–Picture Vocabulary			
Raw score	15.8 (2.9)	15.7 (3.0)	16.9 (2.4)
W score	469.4 (11.4)	469.0 (11.6)	473.5 (8.1)
Catts Deletion Task raw score	5.4 (7.1)	5.1 (7.0)	8.6 (7.8)
SSRS–Externalizing behavior	0.4 (0.5)	0.4 (0.5)	0.6 (0.5)
SSRS–Internalizing behavior	0.4 (0.3)	0.4 (0.3)	0.4 (0.3)
CBQ–Effortful control	4.9 (0.8)	4.9 (0.8)	4.4 (0.8)

Note. CELF-4 = Clinical Examination of Language Fundamentals–Fourth Edition; CFD = Concepts and Following Directions; WJ-III = Woodcock-Johnson–III Test of Cognitive Abilities; SSRS = Social Skills Rating System; CBQ = Child Behavior Questionnaire.

[a]Typical therapy schedule subsample consists of children who received between 15 and 80 therapy sessions over a span of 30 to 40 weeks (0.5–2.4 sessions per week). Shortened therapy schedule subsample consists of children who received fewer than 30 weeks of therapy. There are also three children who were exposed to extremely intensive treatment schedules (more than 100 sessions, 3–4 sessions per week), but the demographic statistics of this subgroup are not reported here due to the limited sample size.

behavior and SSRS–Internalizing behavior, were used to measure the level of problem behavior the children displayed. Upon entry to the study, caregivers completed SSRS and rated various aspects of the children's behavior on a 3-point scale (0 = *never,* 1 = *sometimes,* 2 = *very often*). Among the items, six items measure the children's externalizing behavior, and six items measure their internalizing behavior. Two composite scores (external, internal) were computed as the mean item responses, where larger values indicate higher ratings of problem behavior.

Self-Regulation

The Children's Behavior Questionnaire–Very Short Form (CBQ; Putnam & Rothbart, 2006) was completed by caregivers in the fall of the year at the same time they completed the SSRS. Caregivers responded to 36 statements about their children on a 7-point scale with descriptive anchors (e.g., 1 = *extremely untrue,* 7 = *extremely true*). The tool captures three subscales (surgency, negative affect, and effortful control), of which one subscale, effortful control, was used in this study, consisting of the mean score

Table 3. Descriptive data for children's treatment intensity over academic year.

Characteristic	All participants	Subsample[a]	
		Typical therapy	Shortened therapy
N	272	251	42
Total number of weeks of therapy	34.0 (6.4)	35.6 (2.0)	15.5 (9.3)
Total number of sessions received	43.9 (17.8)	45.3 (13.7)	17.2 (10.0)
Average time (min) spent on language targets per session	11.8 (4.7)	11.8 (4.7)	
Cumulative time (min) spent on language targets	537.2 (275.4)	528.5 (261.3)	

[a]Typical therapy schedule subsample consists of children who received between 15 and 80 therapy sessions over a span of 30 to 40 weeks (0.5–2.4 sessions per week). Shortened therapy schedule subsample consists of children who received fewer than 30 weeks of therapy. There are also three children who were exposed to extremely intensive treatment schedules (more than 100 sessions, 3–4 sessions per week), but the demographic statistics of this subgroup are not reported here due to the limited sample size.

for 12 items. Effortful control (one's ability to inhibit or override dominant tendencies) is a major contributor to one's self-regulation (Ponitz et al., 2009), and thus the effortful control subscale of the CBQ was used in this study to measure this construct. The CBQ Very Short Form has adequate criterion validity, internal consistency, and cross-informant agreement (Putnam & Rothbart, 2006). Children in the present study received a mean rating of 4.9 (*SD* = 0.8) on the effortful control subscale.

Treatment Intensity

Treatment intensity generally refers to the amount of treatment provided and/or received by a child. Though the focus of this study was not on treatment intensity, it was necessary to control for the different amounts or levels of therapy each child received, as there was considerable variability among children in how often they were seen by their SLP. By controlling for variability in children's treatment intensity, we were able to focus this study on understanding how individual differences among children, when controlling for the amount of treatment received, are related to children's language gains over an academic year. To represent treatment intensity in this study as experienced by individual children, we captured three measures: (a) frequency of sessions, (b) average length of language-focused intervention across sessions, and (c) cumulative time in language-focused intervention across sessions. Table 3 provides descriptive data for these three parameters for the study sample.

Frequency of sessions. The frequency of sessions was calculated on the basis of weekly therapy logs (i.e., one log for each week of the school year) submitted by each SLP. For this study, treatment frequency was defined as the total number of therapy sessions a child received over the academic year, which averaged 43.9 sessions for the children in this sample (*SD* = 17.8, range 3–154).

Average length of language-focused intervention. The average length of language-focused intervention (dosage) was calculated on the basis of extensive coding of therapy videos submitted by SLPs for each child in the study. In particular, the time SLPs spent targeting any one of nine language-focused targets was captured across three therapy sessions (i.e., one from the fall, winter, and spring) for each participating child, representing detailed analysis of 589 therapy sessions. These nine targets included any therapeutic time spent targeting grammar, vocabulary, listening comprehension, communicative functions, discourse, narration, abstract language, metalinguistics, or literacy, as detailed in a comprehensive coding scheme (Language Intervention Observation Scale [LIOS]; Justice & Schmitt, 2010). LIOS coding was conducted using the Noldus Observer XT software program, which allowed for moment-by-moment coding of each therapy session in its entirety. LIOS coders completed over 40 hr of training and were required to achieve reliability criteria on master-coded therapy sessions (.70 kappa overall). On average, language was targeted 11.8 min per session (*SD* = 4.7, range 0.9–22.7). The procedures used to code therapy videos are detailed elsewhere (Tambyraja, Schmitt, Justice, Logan, & Schwarz, 2014).

Cumulative time in language-focused intervention. The cumulative time in language-focused intervention, representing the absolute value of time over the year in which children received language-focused intervention, was the multiplication of the aforementioned two parameters, namely, the number of therapy sessions received by each child and the average amount of time spent on language-focused intervention across three therapy sessions for each participating child. Cumulative intensity reported for this sample averaged 537.2 min over the academic year (*SD* = 275.4, range 28.3–1,644.9).

Missing Data

All children in the sample had complete data for the language skill measure (i.e., CELF-4 test scores), both at baseline and at posttest. For the child-level direct measures of nonverbal cognition, memory, vocabulary, and phonological awareness, all children had complete data. However, there were some missing data for the family background variables (e.g., annual household income) and indirect report assessments completed by caregivers (i.e., SSRS and CBQ). For these measures, about 20%–25% of the data were missing due to nonresponse to questionnaires. Also, whereas the frequency of treatment was fully observed, 20% of treatment dosage (i.e., the average length of language-focused intervention) data were missing.

To address the missing data in the predictor variables, we applied multiple imputation (Little & Rubin, 1987) to treat missing data instead of using listwise deletion, which has been shown to produce biased results and low power (Graham, 2012). Inclusive imputation (Schafer & Olsen, 1998) was conducted, where multiple imputation models included all outcome measures as well as other variables theoretically or empirically related to the outcomes or rate of missingness. To account for the nested nature of the data, a multilevel imputation model was applied by treating SLP as a random component. Twenty data sets (Enders, 2010) were imputed and fitted to a multilevel regression model using the multiple imputation module in Mplus 7.11 (Muthén & Muthén, 2006).

Results

Preliminary Analyses

The second column ("All participants") of Table 2 describes the sampled children's average levels in language skills and other relevant characteristics during the fall of the academic year. With respect to the children's language skills, the sample averaged about 2 *SD* below the mean on the CELF-4 Core Language Composite standard score (*M* = 68.8, *SD* = 17.3). Nonverbal cognition was generally higher (*M* = 88.4, *SD* = 12.0), although there was considerable variability. Table 4 provides the mean gains in language skill observed for the overall sample of children, as

Table 4. CELF gain scores from pretest to posttest for participants and subsamples.

CELF score	All participants (*n* = 272)			Typical therapy schedule (*n* = 247)			Shortened therapy schedule (*n* = 22)		
	M	*SD*	Range	*M*	*SD*	Range	*M*	*SD*	Range
CELF-CFD gain score	5.41	6.60	–21 to 31	5.44	6.58	–21 to 31	4.90	7.12	–7 to 24
CELF-WS gain score	3.11	4.14	–9 to 17	3.17	4.14	–9 to 17	2.68	4.20	–7 to 10
CELF-RS gain score	5.44	6.63	–28 to 28	5.60	6.62	–28 to 28	3.77	6.85	–12 to 15
CELF-FS gain score	5.47	7.25	–22 to 29	5.67	7.14	–22 to 29	3.86	8.19	–9 to 21
CELF sum gain	19.44	13.26	–18 to 67	19.87	13.43	–18 to 67	15.23	10.58	–12 to 36

Note. CELF = Clinical Examination of Language Fundamentals–Fourth Edition (CELF-4); CFD = Concepts and Following Directions; WS = Word Structure; RS = Recalling Sentences; FS = Formulated Sentences.

well as two subsamples of children created upon further inspection of the therapy intensity variables.

In particular, a subsample of students was identified who received therapy that was outside of the range we would expect. Here, we highlight the observed differences. Although the majority of the participants (n = 247, 91%) had between 15 and 80 therapy sessions over a span of 30 to 40 weeks ("Typical schedule"), approximately 8% (n = 22) of the sampled children received fewer than 30 weeks of treatment and a limited number of sessions ("Shortened schedule"). The rest of the sample (n = 3, 1%) had more than 100 sessions in 40 weeks ("Intensive schedule"). Figure 1 shows a serious negative skew in the number of weeks of therapy provided (upper left panel), and a serious positive skew in the number of sessions delivered (upper right panel). For the 22 children with shortened schedules, due to the incompleteness of their therapy videos, measures of treatment dosage were not available.

With an adjusted alpha level of 0.006 (i.e., 0.05 divided by eight pretreatment measures), children with shortened schedules scored significantly higher than those with typical schedules on the baseline language tests (CELF-4 total standard score, diff = 11.6; $p < .001$). Comparison of other characteristics between the two subsamples, although not statistically significant, showed a tendency for children with shortened schedules to achieve higher scores on the cognitive and literacy measures (memory: diff = 15.4, p = .032; vocabulary: diff = 4.4, p = .025; phonological awareness: diff = 3.5, p = .057), but lower scores on the behavioral measures (effortful control: diff = 0.51, p = .025). Detailed descriptive statistics of pretreatment characteristics for the subsamples can be found in Tables 1 and 2.

Considering that the subgroups differed significantly in their language skills at baseline, and that they received differing intensities of therapy, we decided to conduct the primary analyses first on the basis of the full sample, and then on the predominant subsample (i.e., children with typical treatment schedule). The analyses were not replicated with the other subsample due to sample size limitation.

Language Gains for the Sample

An initial aim of this study was to examine gains in language skills among children with LI within the public schools. To address this aim, descriptive statistics were used to measure the change in CELF scores from fall (baseline, or pretest) to spring (follow-up, or posttest). For the 272 participants, CELF total scores (i.e., the sum of raw scores of the four subtests) increased from 61.2 (*SD* = 34.0) to 80.7 (*SD* = 36.6), corresponding to 19.5 points, or approximately 0.55 *SD* of the baseline level, and each of the four subtests also showed increases (Tables 2 and 4). The distribution of the total CELF gain scores (calculated as the fall score subtracted from the spring score; see Figure 2, upper graph) showed that the majority of the participants' scores increased by 10 to 30 points. The standard deviation of gain was 13.3 points, and the range of change was 85 (from −18 to +67), indicating that the gains varied substantially among children.

Considering the two subsamples of children described in the previous section, those with typical schedules witnessed somewhat larger gains in language skills than those with shortened schedules, as measured by the raw scores of all four subtests of CELF as well as the composite score (see Table 4). The differences, however, were not statistically significant. The lower graph of Figure 2 shows that the children with higher scores on the CELF were disproportionately more likely to be from the typical therapy schedule group.

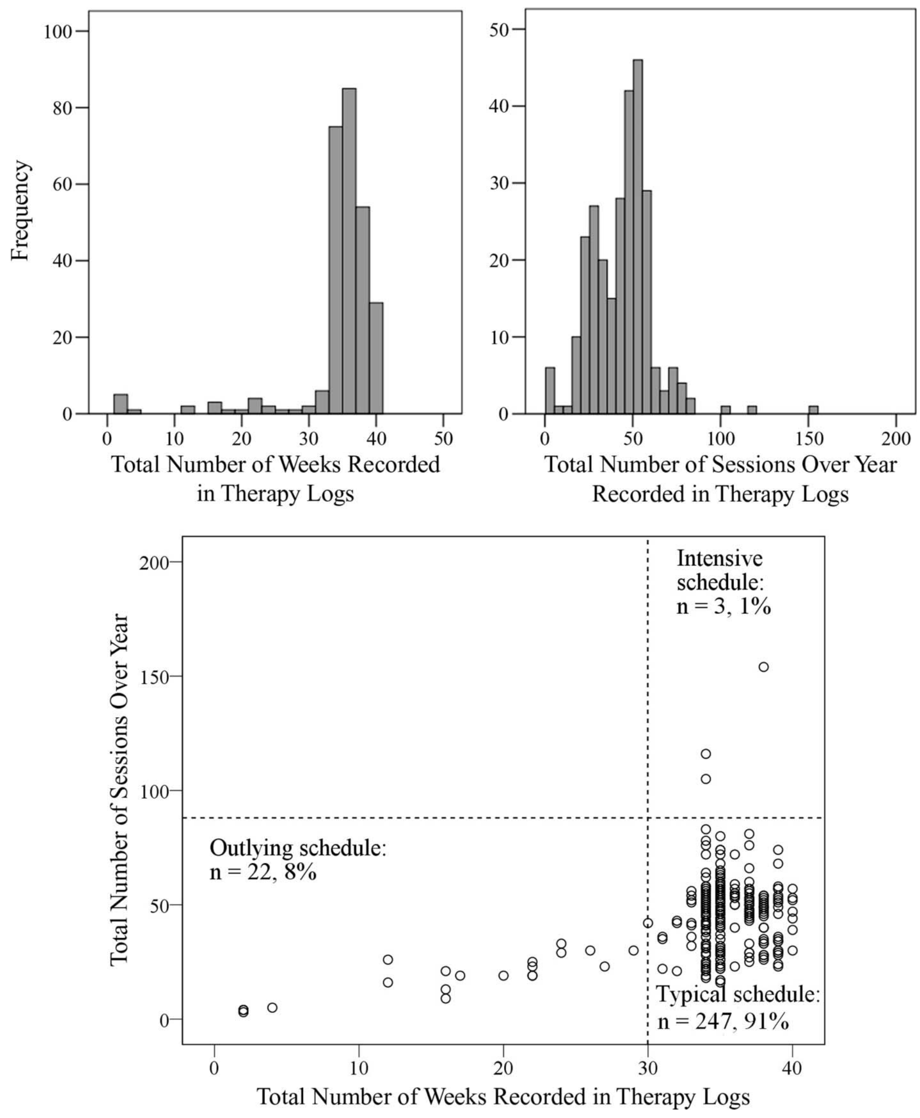
Predictors of Language Gains for Children With LI

Unconditional Model

Given the large amount of variation in the change of language skills among children with LI, regression models were fit to the data to predict children's language gains over the academic year from the selected child-level characteristics, controlling for children's baseline language score, treatment-intensity parameters, and family background variables. Given the nested structure of the data (i.e., children were nested within SLPs), the variation among SLPs was accounted for by multilevel models using Mplus 7.11 (Muthén & Muthén, 2006). Rather than identifying one estimate of the intercept and slope that best fit all points, multilevel models allowed a separate line of best fit to be estimated for each SLP in the sample.

The outcome of interest was children's language gains over the academic year, as measured by the change

Figure 1. Histograms displaying the distribution of the number of weeks of therapy (upper left) and number of therapy sessions (upper right), and scatter plot displaying the relationship between the number of weeks and number of sessions (lower graph).



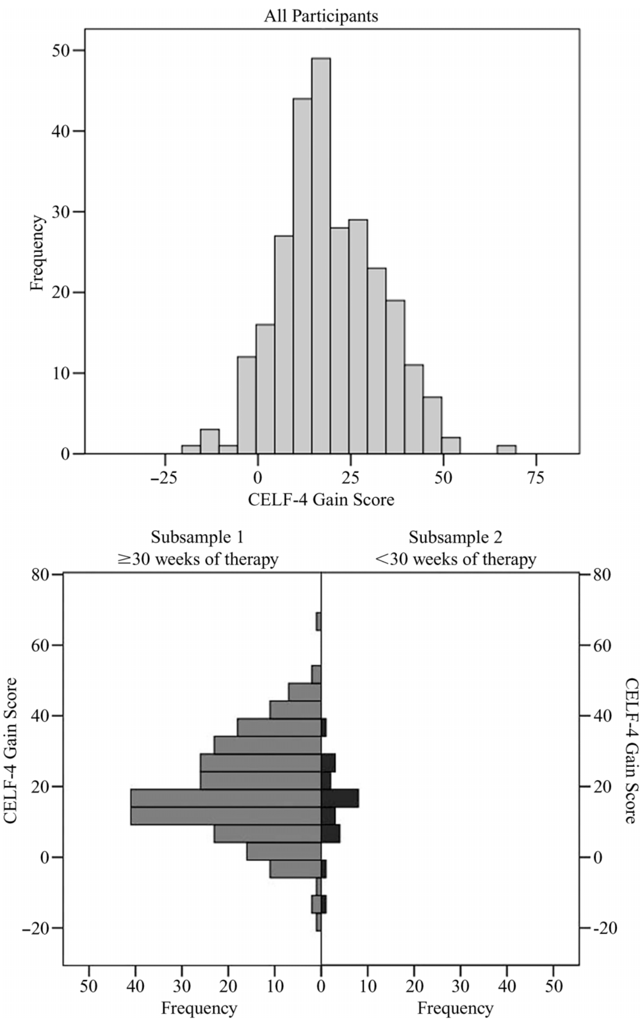
in CELF composite score from fall to spring. Without inclusion of any predictors, the unconditional intraclass correlation (variance attributable to SLP divided by total variance) was 0.09 for the whole sample, indicating that 9% of variance in CELF gain scores was attributable to differences between SLPs, and the remaining 91% of variance was due to individual differences between children.

Conditional Model

To identify characteristics of children that are significant predictors of gains in language scores, we next added predictors to the model, including children's baseline language scores (as measured by CELF-4 in the fall), family background variables (maternal level of education and family income levels), and the three treatment intensity parameters. After controlling for these variables, we included the eight child-level characteristics of interest, namely, age in months, nonverbal cognition (KBIT raw score), memory (CELF-4 Number Memory raw score), phonological awareness (Catts Deletion Task raw score), vocabulary (WJ-III–Picture Vocabulary raw score), externalizing problem behavior (SSRS–Externalizing subscale average rating), internalizing problem behavior (SSRS–Internalizing subscale average rating), and effortful control (CBQ Effortful Control average rating). Table 5 displays the bivariate Pearson correlation coefficients for all of the variables in the model.

As shown in Table 6, for the overall sample (n = 272), after controlling for baseline language (b = −0.18, β = −0.47,

Figure 2. Histograms displaying the distribution of the change in the Clinical Examination of Language Fundamentals–Fourth Edition (CELF-4) composite score for all participants (upper graph) and side-by-side comparison of two subsamples (lower graph).



$p < .001$), maternal education ($p = .562$), family income ($p = .631$), and the three intensity parameters (frequency, b = 0.12, β = 0.16, $p = 0.276$; dosage, b = 0.28, β = 0.11, $p = 0.449$; cumulative intensity, b = −0.01, β = −0.20, $p = 0.24$), two characteristics appeared to be highly predictive of the children's language gains over time, namely, phonological awareness (b = 0.57, β = 0.32, $p = .001$) and vocabulary (b = 1.22, β = 0.28, $p = .001$). In particular, one point of increase on the phonological awareness measure corresponded with an expected increase of 0.57 points in the CELF gain score, and one point of increase on the vocabulary measure corresponded with 1.22 extra points in the gain score. A combination of over 20 imputed data sets using procedures outlined in Harel (2009) shows that the within-level R^2 is 0.151, indicating that 15.1% of the child-level variance is accounted for by the model. To test the robustness of these findings, the same multilevel regression model was fitted to the typical schedule subsample ($n = 247$), omitting the children with outlying schedules, and the pattern of results was replicated.

Beyond examining only the statistical significance of the predictors investigated, we also sought to evaluate the relative importance of the predictors of interest. As is commonly noted in the literature (Harel, 2009; Pratt, 1987), the relative importance of a predictor reflects how much it contributes to the prediction of an outcome variable in the presence of the other correlated predictors. Although the standardized coefficients (β) are traditionally used to measure predictor importance, they are highly unstable with the addition or removal of extra predictors, and they lack the additive property necessary for evaluating predictor contribution (i.e., they do not add up to 1; Azen & Budescu, 2009). Numerous other approaches have been proposed as alternatives to measure relative importance (Budescu, 1993; Pratt, 1987; Thomas, Hughes, & Zumbo, 1998), although the extension of those methods to multilevel framework has been relatively scarce and has appeared in the literature only recently. For the current study, we used the Pratt index adapted for hierarchical linear modeling (HLM; Liu, Zumbo, & Wu, 2014) to evaluate the predictors' importance. In a random-intercept multilevel model, the index is simply calculated as

$$d_i = \beta \times r_i / R^2,$$

where β is the standardized β coefficient, r_i is the Pearson correlation between a predictor and the outcome, and R^2 is the within-level R^2 value. As an R^2-based statistic, the Pratt index measures the percentage of variance accounted for by each predictor uniquely. It is worth noting that a negative Pratt value is possible due to the uncertainty of estimates. Therefore, we further computed a 90% confidence interval for each Pratt index (Thomas, Zhu, & Decady, 2007) to capture the amount of uncertainty in predictor importance.

Pratt indices for the full sample are displayed in the last two columns of Table 6. Thomas (1992) suggested that predictors with indices below 1/(2 × number of predictors) can be considered unimportant. Following this rule of thumb, five predictors appeared to be potentially important (Pratt index > .038): phonological awareness (.351), vocabulary (.324), externalizing behavior (.091), nonverbal cognition (.088), and child age (.077). The relative contribution of each predictor to the child-level variance can be computed as the product of Pratt index and within-level R^2 (see Figure 3). Among these predictors, phonological awareness uniquely accounted for 35.1% of the explained variance (95% confidence interval = 5.7%, 64.4%) or 5.3% of the total child-level variance in language gain, whereas vocabulary accounted for 32.4% of the explained variance (95% confidence interval = 5.3%, 59.6%), or 4.9% of the total variance. Externalizing behavior, nonverbal cognition, and age accounted for over 25% of the explained variance combined (or 3.8% of the total variance). However, these three individual contributions did not differ significantly from zero.

Table 5. Matrix of bivariate correlation coefficients (Pearson *r*) of key variables.

Characteristic	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. CELF gain score	—														
2. CELF baseline score	.006	—													
3. Mother having bachelor's degree	.008	.102	—												
4. Low income (≤ $40K)	−.045	−.160	−.444	—											
5. Number of sessions	.021	−.067	.103	−.041	—										
6. Language-focused intervention	.024	−.107	−.036	.033	−.158	—									
7. Cumulative intensity	−.011	−.078	.034	−.005	.552	.690	—								
8. Age in months	.122	.361	−.030	.020	.051	.126	.104	—							
9. Nonverbal cognition	.103	.509	.030	−.014	−.024	−.040	.021	.265	—						
10. Memory	.052	.713	.082	−.075	−.082	−.049	−.042	.250	.527	—					
11. Phonological awareness	.168	.696	.022	−.117	−.074	.019	.036	.356	.434	.652	—				
12. Vocabulary	.176	.625	.179	−.176	−.053	−.058	−.060	.223	.362	.490	.431	—			
13. Externalizing behavior	−.115	−.085	−.094	.189	−.140	.050	−.048	−.054	.042	−.054	−.080	−.037	—		
14. Internalizing behavior	−.097	−.235	.019	.110	.020	.095	.074	−.091	−.075	−.171	−.129	−.150	.625	—	
15. Self-regulation	−.008	.063	.041	.115	.132	−.128	−.065	−.162	.010	.018	.016	−.005	−.087	−.001	—

Note. CELF = Clinical Examination of Language Fundamentals–Fourth Edition (CELF-4); Nonverbal cognition = scores on the Kaufman Brief Intelligence Test; Memory = scores on the Numbers Repetition subtest from the CELF-4; Phonological awareness = scores on the Catts Deletion Task; Vocabulary = scores on the Picture Vocabulary subset of the Woodcock-Johnson III Test of Cognitive Abilities; Externalizing and internalizing behavior = scores from the Social Skills Rating System; Self-regulation = scores on the effortful control subscale of the Child Behavior Questionnaire.

For the typical therapy schedule group, the potentially important predictors (Pratt > .038) included: phonological awareness (.425), vocabulary (.330), externalizing behavior (.067), nonverbal cognition (.060), treatment dosage (.052), cumulative intensity (.046), and age (.041). With this subsample, the unique contribution of phonological awareness became especially salient (42.5% of the explained variance, and 7.1% of the total child-level variance), whereas the contribution of vocabulary was similar to that of the full sample (33.0% of the explained variance, and 5.5% of the total variance). Contrary to the full sample, where treatment intensity parameters contributed little to the variation of the outcome, in this subsample, treatment dosage and cumulative intensity accounted for 9.8% of the explained variance (or 1.6% of the total variance) in language gains, although their

Table 6. Predicting language gains: Results of multilevel random-intercept regression models (all participants, *n* = 272).

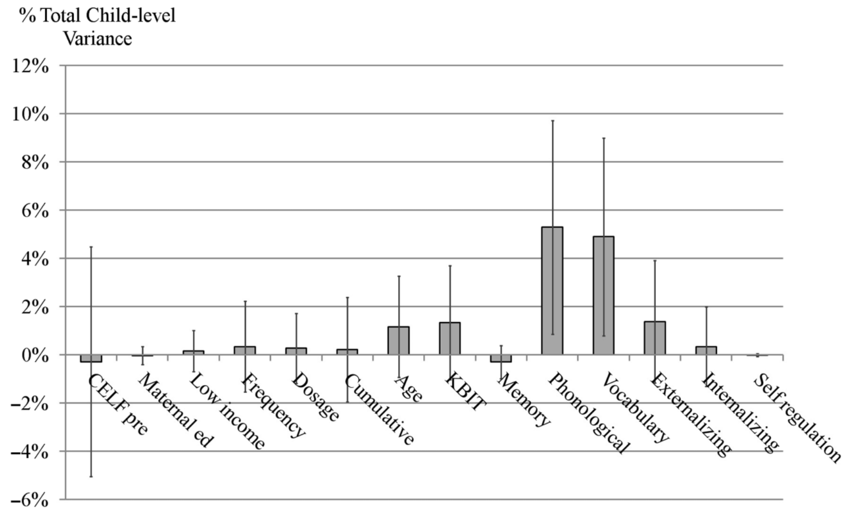
Characteristic	Estimate	β[a]	*SE*	*p*	Pratt index	90% confidence interval of Pratt index
Baseline language	−0.18	−0.47	0.11	< .001***	−.019	(−.335, .297)
Mother having bachelor's degree	−1.19	−0.04	0.07	.562	.002	(−.027, .022)
Low income (≤ $40K)	−0.92	−0.04	0.07	.631	.011	(−.046, .067)
Frequency: Number of sessions	0.12	0.16	0.15	.276	.022	(−.102, .147)
Dosage: Language-targeted instruction per session	0.28	0.11	0.15	.449	.018	(−.078, .113)
Cumulative time of language-targeted instruction	−0.01	−0.20	0.17	.240	.014	(−.129, .157)
Age in months	0.15	0.10	0.07	.144	.077	(−.062, .217)
Nonverbal cognition	0.39	0.13	0.08	.115	.088	(−.069, .244)
Memory	−0.28	−0.06	0.10	.550	−.020	(−.066, .026)
Phonological awareness	0.57	0.32	0.09	**.001****	**.351**	**(.057, .644)**
Vocabulary	1.22	0.28	0.08	**.001****	**.324**	**(.053, .596)**
Externalizing behavior	−3.26	−0.12	0.09	.192	.091	(−.078, .259)
Internalizing behavior	−1.26	−0.04	0.09	.695	.022	(−.087, .132)
Self-regulation	0.06	0.00	0.06	.950	.000	(−.004, .004)
Intraclass correlation (ICC)	0.090					
Within-level R^2	0.151					

Note. Regression results are combined over 20 imputed data sets. CELF = Clinical Examination of Language Fundamentals–Fourth Edition (CELF-4); Nonverbal cognition = scores on the Kaufman Brief Intelligence Test; Memory = scores on the Numbers Repetition subtest from the CELF-4; Phonological awareness = scores on the Catts Deletion Task; Vocabulary = scores on the Picture Vocabulary subset of the Woodcock-Johnson III Test of Cognitive Abilities; Externalizing and internalizing behavior = scores from the Social Skills Rating System; Self-regulation = scores on the effortful control subscale of the Child Behavior Questionnaire.

[a]Standardized coefficients are indicated by beta (β).

**$p < .01$. ***$p < .001$.

Figure 3. Percentage of total child-level variance in language gains accounted for by pretreatment characteristics (error bars representing 90% confidence intervals), where CELF = Clinical Examination of Language Fundamentals–Fourth Edition (CELF-4), and KBIT = Kaufman Brief Intelligence Test.



unique contributions were not significantly different from zero.

Discussion

This study serves to advance our understanding of the characteristics of children with LI that can explain their gains in language skills over an academic year within the public schools. Involving a sample of nearly 300 clinically identified primary-grade pupils with LI, this study helps to identify characteristics of children that may serve as prognostic indicators of their language growth over time, namely, their phonological awareness and vocabulary skill. However, because this work is exploratory, we focus our discussion not only on those predictors that demonstrated statistical significance, but also those variables that show relative importance in the presence of the other correlated predictors (those predictors that had high Pratt indices).

To this end, study findings showed that four cognitive and noncognitive characteristics, as well as children's age, were potentially important for predicting children's language gains over time. In order of importance, children's phonological awareness, vocabulary, externalizing behavior, and nonverbal cognition were each unique predictors of variance in children's language gains during an academic year in which they received school-based treatment. Of these, phonological awareness and vocabulary offered the greatest, most substantial predictive power in understanding variability in gains: Children with higher levels of phonological awareness and vocabulary at the start of the academic year exhibited the greatest gains in language skill over time. This and additional findings of interest are discussed here, as well as ways in which this work can contribute to clinical practice and future treatment studies for children with LI.

Children's phonological awareness and vocabulary skill together accounted for nearly 70% of the explained variance in children's language gains over an academic year of schooling. Put simply, children who had relatively high levels of oral-language skill, on the basis of their proficiency on phonological awareness and vocabulary assessments implemented in fall of the year, made the greatest gains over the academic year. This finding, although not entirely unexpected because of some prior evidence involving typically developing children receiving language-focused interventions (Justice, Meier, & Walpole, 2005; Lawrence, Givens, Branum-Martin, White, & Snow, 2014; Penno et al., 2002), suggests that a Matthew effect may characterize the language development of children with LI.

The Matthew effect refers to the phenomenon in which "the rich get rich, while the poor get poorer," and it has frequently been observed to describe patterns in children's growth trajectories, particularly in the area of reading achievement (Pfost, Hattie, Dörfler, & Artelt, 2013). The Matthew effect refers to the observation that the gap between children with higher and lower skills tends to grow larger with time. For instance, in an application to children with LI, Morgan et al. presented evidence of a Matthew effect in modeling reading trajectories from kindergarten to fifth grade for children with LI as compared with peers without disabilities (Morgan et al., 2011). At kindergarten, the peer group significantly outperformed the students with LI in reading achievement, but the gap was modest (about 5 score points). However, by fifth grade, the achievement gap had widened, differentiating the two groups by more than 20 points in a poor-get-poorer phenomenon. This growing gap between groups of pupils is often referred to as a Matthew effect.

The present study suggests that Matthew effects may be operating within the context of treatment provision within the public schools for children with LI in the early primary grades. In particular, children with relatively higher levels of language skill at the start of the academic year make greater gains than children with relatively lower levels of skill, using measures of phonological awareness and vocabulary skill. Although the research design used here is not experimental, the results are similar to those of experimental studies in which children with higher levels of pretreatment language skill benefit more from language-focused interventions than children with lower skill levels (Johanson, Justice, & Logan, 2015; Justice et al., 2010; Penno et al., 2002). For instance, Penno et al. implemented a language intervention designed to improve young children's vocabulary knowledge. Children with higher levels of vocabulary skill at the start of the intervention gained significantly more than children with lower levels of vocabulary skill. Such circumstances create a paradox of sorts, in that interventions serve to provide greater benefit to those who presumably need the intervention less.

The results of this study suggest the need to explore the presence of Matthew effects for children with LI further, including the mechanisms through which the Matthew effects are exerted. For instance, the Matthew effect might result from SLP-specific factors, such as a tendency to provide less-robust language therapy to children with less-developed language skills; in turn, these children may gain less from therapy over time. On the other hand, the effect might reflect child-specific factors. Children with low levels of phonological awareness and/or vocabulary skill may exhibit processing limitations that hinder language development (Rice, Buhr, & Nemeth, 1990). In turn, such processing limitations would be associated with relatively slow gains in language skill over time. At the same time, it is also relevant that the children in this study varied substantially in the severity of their LI. Severity of LI may interact with other child-specific factors to affect language growth during speech-language intervention. For instance, children with severe language problems and phonological processing issues may be especially vulnerable for lags in language growth. The findings presented here point to the need for researchers to further investigate the possibility of a Matthew effect and to explore the mechanisms that might explain it.

The second finding of interest is the modest but possibly meaningful effects of children's externalizing behavior and nonverbal cognition on their language gains over time. Although the effects of these predictors were modest, the study results suggested that children with lower levels of problem behaviors and higher levels of nonverbal cognition exhibited greater gains in language skill during an academic year as compared with children not exhibiting these characteristics. Similar results have been reported for other populations of children with language disorders. Paul et al. followed children with autism from 2 to 4 years of age and found that children with better language outcomes at age 4 years had higher nonverbal cognition scores at age 2 years (Paul, Chawarska, Cicchetti, & Volkmar, 2008). Beitchman et al. showed an overlap in behavioral problems and severe cases of LI, such that children with significant LI at school entry were likely to exhibit behavioral problems over time (Beitchman et al., 1996). Our work indicates that higher levels of behavioral problems may be associated with less gain in language skill over time for children with LI in the public schools. To this end, both nonverbal cognition and behavioral problems should be explored as potential prognostic indicators of language development for children with LI.

The findings of this study should also be considered with respect to predictor variables that did not appear influential to understanding variability in language gains for children with LI. In particular, children's working memory, internalizing behavior, and self-regulation did not appear to be meaningful predictors of change. This is somewhat surprising, because studies of children within classroom contexts have described these as important "learning-related behaviors" that serve to create conditions under which children can effectively learn (McClelland, Acock, & Morrison, 2006; Stipek, Newton, & Chudgar, 2010). For instance, children with high levels of self-regulation are able to control their own impulses and manage distractions within classroom activities, thereby facilitating their learning within this environment. However, for the children in this study who were receiving therapy from school-based SLPs, these characteristics were not influential to their gains in language skill. Perhaps this is due to the treatment contexts experienced by most of the children, which typically featured one-on-one and small-group interactions, as there are fewer distractions to students and greater support available from the SLP, given the smaller adult-child ratio as compared with classroom settings.

There are several implications that can be drawn from this work with respect to both research and practice. With respect to research, some studies suggest that intervention for children with LI often has modest or even equivocal effects on children's language growth (Cirrin & Gillam, 2008; Law, Garrett, & Nye, 2004). However, it is possible that the effects of language intervention are dependent upon characteristics of the children being served. The results of the present study may be helpful for pinpointing those moderators warranting attention as possible moderators within treatment research. For instance, given that children's phonological awareness appears to explain a considerable amount of variance in children's language gains over time, it would be reasonable to consider whether the effects of specific language-focused intervention are conditional on children's phonological awareness, as well as vocabulary, externalizing behavior, and nonverbal cognition.

With respect to practice, the findings may be helpful for clinicians as they seek to understand prognostic indicators for the children they are serving. Clinicians may have certain students who appear to be inadequately progressing within treatment, and they may wish to consider whether the results of this study are helpful to understanding this lack of response. For instance, do the nonresponders have very low levels of phonological awareness or vocabulary skill? However, the results of this study do not speak to what the clinician should do, and thus we can only speculate. One possibility is to focus intensively on building those skills (phonological awareness and vocabulary) that seem to contribute to language gain over time. Nonetheless, it is important to recognize that the correlation between phonological awareness or vocabulary and children's language gains over time does not imply causation. Thus, improving children's skills in these areas will not necessarily result in elevated treatment outcomes.

Several limitations to this study warrant attention. The first limitation is the correlational nature of this work. The relations among the variables investigated in this work are not necessarily causal in nature, and thus they should not be interpreted as such. The second limitation concerns the clinical sample of children with LI. Although the sample is heterogeneous as compared with other large-scale studies of children with LI (e.g., Catts et al., 2002; Gillam et al., 2008), it is more homogeneous than the children with LI served by SLPs in the public schools. For instance, our ascertainment procedures eliminated children for whom language problems were not their primary complaint, children who did not speak English at home, and children without caregiver consent. Thus, the generalizability of our results to clinical caseloads is not clear. Third, several measures of children's pretreatment characteristics relied on parent report, including the social-skills assessment and self-regulation measure. Although parent report is often used in research on children's social skills and self-regulation, it may not adequately represent these skills outside of the home environment (Dinnebeil et al., 2013). For instance, children's ability to regulate their own behavior or to work independently on tasks may be context-dependent. Therefore, replication of the present results with assessments of children's skills in the school and therapeutic context is recommended. Last, we also point out that our measure of language gain represented a composite of various linguistic domains (e.g., grammar, vocabulary) and modalities (i.e., receptive and expressive skill). Future work should explore whether the predictors identified as important in the present study are observed when other outcomes are considered, such as receptive and expressive skill treated separately. At the same time, we must also note that children's therapeutic experiences were not mapped to their language growth; that is, we did not map children's language goals targeted in therapy to their change over time. To fully understand children's language growth during therapy, it is important that future research look more intensively at the nature of therapy itself in relation to children's change.

To close, there is a significant need for research on how to maximize language gains for children with LI. In addition to investigations of experimental treatments, there is a need to explore child-level cognitive and noncognitive characteristics that may explain variability in their language gains over time. These are important for understanding the nature of LI among school-aged children, but also as possible prognostic indicators of treatment response and as potential moderators of treatment exposure. The results of the present study serve to identify two child-level characteristics that uniquely predict language gains over an academic year for children with LI: phonological awareness and vocabulary skill. Researchers should explore how these characteristics may serve to moderate treatment outcomes for children, whereas clinicians should assess how these characteristics may factor into understanding treatment response among children undergoing therapy.

Acknowledgments

The authors are grateful to the numerous speech-language pathologists, families, and children who participated in the original research utilized in this study. The original research was supported by Grant R324A090012 from the U.S. Department of Education, Institute of Education Sciences, to Laura M. Justice.

References

Archibald, L. M., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders, 41*(6), 675–693.

Azen, R., & Budescu, D. V. (2009). Applications of multiple regression in psychological research. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology* (pp. 285–310). Thousand Oaks, CA: Sage.

Beitchman, J. H., Wilson, B., Brownlie, E., Walters, H., Inglis, A., & Lancee, W. (1996). Long-term consistency in speech/language profiles: II. Behavioral, emotional, and social outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry, 35*(6), 815–825.

Bishop, D. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language & Communication Disorders, 49*(4), 381–415.

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE Consortium. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLoS ONE, 11*(7), e0158753. https://doi.org/10.1371/journal.pone.0158753

Briscoe, J., Bishop, D. V., & Frazier Norbury, C. (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *Journal of Child Psychology and Psychiatry, 42*(3), 329–340.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin, 114*(3), 542–551. https://doi.org/10.1037/0033-2909.114.3.542

Camarata, S. M., Nelson, K. E., & Camarata, M. N. (1994). Comparison of conversational-recasting and imitative procedures for training grammatical structures in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 37*(6), 1414–1423.

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*(6), 1142–1157. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12546484 or http://jslhr.pubs.asha.org/data/Journals/JSLHR/929267/jslhr_45_6_1142.pdf

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*(1), 38–50. https://doi.org/10.1044/0161-1461(2001/004)

Cirrin, F. M., & Gillam, R. B. (2008). Language intervention practices for school-age children with spoken language disorders: A systematic review. *Language, Speech, and Hearing Services in Schools, 39*(1), S110–S137. https://doi.org/10.1044/0161-1461(2008/012)

Cole, K. N., Mills, P. E., & Kelley, D. (1994). Agreement of assessment profiles used in cognitive referencing. *Language, Speech, and Hearing Services in Schools, 25*(1), 25–31.

Conti-Ramsden, G., & Botting, N. (1999). Classification of children with specific language impairment: Longitudinal considerations. *Journal of Speech, Language, and Hearing Research, 42*(5), 1195–1204.

Côté, I., Rouleau, N., & Macoir, J. (2014). New word acquisition in children: Examining the contribution of verbal short-term memory to lexical and semantic levels of learning. *Applied Cognitive Psychology, 28*(1), 104–114.

Dance, K. A., & Neufeld, R. W. (1988). Aptitude-treatment interaction research in the clinical setting: A review of attempts to dispel the "patient uniformity" myth. *Psychological Bulletin, 104*(2), 192.

Dinnebeil, L. A., Sawyer, B. E., Logan, J., Dynia, J. M., Cancio, E., & Justice, L. M. (2013). Influences on the congruence between parents' and teachers' ratings of young children's social skills and problem behaviors. *Early Childhood Research Quarterly, 28*(1), 144–152.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Publications.

Farquharson, K., Tambyraja, S. R., Logan, J., Justice, L. M., & Schmitt, M. B. (2015). Using hierarchical linear modeling to examine how individual SLPs differentially contribute to children's language and literacy gains in public schools. *American Journal of Speech-Language Pathology, 24*(3), 504–516.

Gillam, R. B., Loeb, D. F., Hoffman, L. M., Bohman, T., Champlin, C. A., Thibodeau, L., . . . Friel-Patti, S. (2008). The efficacy of Fast ForWord language intervention in school-age children with language impairment: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research, 51*(1), 97–119.

Gillon, G. T. (2002). Follow-up study investigating the benefits of phonological awareness intervention for children with spoken language impairment. *International Journal of Language & Communication Disorders, 37*(4), 381–400.

Graham, J. W. (2012). *Missing data: Analysis and design*. New York, NY: Springer Science & Business Media.

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system: Manual*. Circle Pines, MN: American Guidance Service.

Harel, O. (2009). The estimation of R^2 and adjusted R^2 in incomplete data sets using multiple imputation. *Journal of Applied Statistics, 36*(10), 1109–1118. https://doi.org/10.1080/02664760802553000

Johanson, M., Justice, L. M., & Logan, J. (2015). Kindergarten impacts of a preschool language-focused intervention. *Applied Developmental Science, 20*(2), 1–14.

Justice, L. M., McGinty, A. S., Cabell, S. Q., Kilday, C. R., Knighton, K., & Huffman, G. (2010). Language and literacy curriculum supplement for preschoolers who are academically at risk: A feasibility study. *Language, Speech, and Hearing Services in the Schools, 41*(2), 161–178. https://doi.org/10.1044/0161-1461(2009/08-0058)

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech, and Hearing Services in Schools, 36*(1), 17–32.

Justice, L. M., & Schmitt, M. B. (2010). *Language Intervention Observation Scale (LIOS)*. Columbus, OH: Crane Center for Early Childhood Research and Policy, The Ohio State University.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test–Second Edition*. Circle Pines, MN: American Guidance Service (AGS) Publishing.

Kouri, T. A. (2005). Lexical training through modeling and elicitation procedures with late talkers who have specific language impairment and developmental delays. *Journal of Speech, Language, and Hearing Research, 48*(1), 157–171.

Law, J., Garrett, Z., & Nye, C. (2004). The efficacy of treatment for children with developmental speech and language delay/disorder: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 47*(4), 924–943.

Lawrence, J., Givens, R., Branum-Martin, L., White, C., & Snow, C. (2014). Generating vocabulary knowledge for at-risk middle school readers: Contrasting program effects and growth trajectories. *Journal of Education for Students Placed at Risk, 19*(2), 76–97.

Laws, G., Bates, G., Feuerstein, M., Mason-Apps, E., & White, C. (2012). Peer acceptance of children with language and communication impairments in a mainstream primary school: Associations with type of language difficulty, problem behaviours and a change in placement organization. *Child Language Teaching and Therapy, 28*(1), 73–86.

Little, R. J., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: John Wiley & Sons.

Liu, Y., Zumbo, B. D., & Wu, A. D. (2014). Relative importance of predictors in multilevel modeling. *Journal of Modern Applied Statistical Methods, 13*(1), 2–22.

Majerus, S., & Boukebza, C. (2013). Short-term memory for serial order supports vocabulary development: New evidence from a novel word learning paradigm. *Journal of Experimental Child Psychology, 116*(4), 811–828.

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*(4), 471–490.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*(1), 3.

Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities, 44*(5), 472–488.

Muthén, L. K., & Muthén, B. O. (2006). *Mplus (Version 4.1)*. Los Angeles, CA: Muthén.

Olswang, L. B., & Bain, B. A. (1996). Assessment information for predicting upcoming change in language production. *Journal of Speech, Language, and Hearing Research, 39*(2), 414–423.

Paul, R., Chawarska, K., Cicchetti, D., & Volkmar, F. (2008). Language outcomes of toddlers with autism spectrum disorders: A two year follow-up. *Autism Research, 1*(2), 97–107.

Penno, J. F., Wilkinson, I. A., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*(1), 23.

Petersen, D. B., Gillam, S. L., Spencer, T., & Gillam, R. B. (2010). The effects of literate narrative intervention on children with neurologically based language impairments: An early stage study. *Journal of Speech, Language, and Hearing Research, 53*(4), 961–981.

Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2013). Individual differences in reading development: A review of 25 years of empirical research on Matthew effects in reading. *Review of Educational Research, 84*(2), 203–244.

Ponitz, C. C., McClelland, M. M., Matthews, J., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology, 45*(3), 605.

Pratt, J. W. (1987). Dividing the indivisible: Using simple symmetry to partition variance explained. In T. Pukkila & S. Puntanen (Eds.), *Proceedings of the Second International Conference in Statistics* (pp. 245–260). Tampere, Finland: University of Tampere.

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment, 87*(1), 102–112.

Rice, M. L., Buhr, J. C., & Nemeth, M. (1990). Fast mapping word-learning abilities of language-delayed preschoolers. *Journal of Speech and Hearing Disorders, 55*(1), 33–42. Retrieved from http://jshd.asha.org/cgi/content/abstract/55/1/33

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33*(4), 545–571.

Schmitt, M. B., Logan, J. A., Tambyraja, S., Farquharson, K., & Justice, L. M. (2017). Establishing language benchmarks for children with typically developing language and children with language impairment. *Journal of Speech, Language, and Hearing Research, 60,* 364–378.

Schmitt, M. B., Pentimonti, J. M., & Justice, L. M. (2012). Teacher–child relationships, behavior regulation, and language gain among at-risk preschoolers. *Journal of School Psychology, 50*(5), 681–699.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4).* San Antonio, TX: PyschCorp.

Skibbe, L. E., Grimm, K. J., Stanton-Chapman, T. L., Justice, L. M., Pence, K. L., & Bowles, R. P. (2008). Reading trajectories of children with language difficulties from preschool through fifth grade. *Language, Speech, and Hearing Services in Schools, 39*(4), 475.

Snow, R. E. (1991). Aptitude-treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology, 59*(2), 205.

Stipek, D., Newton, S., & Chudgar, A. (2010). Learning-related behaviors and literacy achievement in elementary school-aged children. *Early Childhood Research Quarterly, 25*(3), 385–395.

Tambyraja, S. R., Schmitt, M. B., Farquharson, K., & Justice, L. M. (2015). Stability of language and literacy profiles of children with language impairment in the public schools. *Journal of Speech, Language, and Hearing Research, 58*(4), 1167–1181.

Tambyraja, S. R., Schmitt, M. B., Justice, L. M., Logan, J. A., & Schwarz, S. (2014). Integration of literacy into speech-language therapy: A descriptive analysis of treatment practices. *Journal of Communication Disorders, 47,* 34–46.

Thomas, D. R. (1992). Interpreting discriminant functions: A data analytic approach. *Multivariate Behavioral Research, 27*(3), 335–362.

Thomas, D. R., Hughes, E., & Zumbo, B. D. (1998). On variable importance in linear regression. *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement, 45,* 253–275.

Thomas, D. R., Zhu, P., & Decady, Y. J., (2007). Point estimates and confidence intervals for variable importance in multiple linear regression. *Journal of Educational and Behavioral Statistics, 32*(1), 61–91.

Throneburg, R. N., Calvert, L. K., Sturm, J. J., Paramboukas, A. A., & Paul, P. J. (2000). A comparison of service delivery models: Effects on curricular vocabulary skills in the school setting. *American Journal of Speech-Language Pathology, 9*(1), 10–20.

U.S. Department of Education. (2015). Table 204.30: Children 3 to 21 years old served under Individuals with Disabilities Education Act (IDEA), Part B, by type of disability: Selected years, 1976–77 through 2013–14. In *Digest of Education Statistics* (2015 ed.). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Washington, K. N., Thomas-Stonell, N., McLeod, S., & Warr-Leeper, G. (2015). Outcomes and predictors in preschoolers with speech-language and/or developmental mobility impairments. *Child Language Teaching and Therapy, 31*(2), 141–157.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Test of Cognitive Abilities*. Itasca, IL: Riverside Publishing.