Published in final edited form as:
Int J Lang Commun Disord. 2013 May ; 48(3): 307–319. doi:10.1111/1460-6984.12008.

Children with Developmental Language Impairment Have Vocabulary Deficits Characterized by Limited Breadth and Depth

Karla K. McGregor, **Jacob Oleson**, **Alison Bahnsen**, and **Dawna Duff**
The University of Iowa

Abstract

Background—Deficient vocabulary is a frequently reported symptom of developmental language impairment but the nature of the deficit and its developmental course are not well documented.

Aims—We aimed to describe the nature of the deficit in terms of breadth and depth of vocabulary knowledge and to determine whether the nature and the extent of the deficit change over the school years.

Methods—A total of 25,681 oral definitions produced by 177 children with developmental language impairment (LI) and 325 grade-mates with normally developing language (ND) in grades 2, 4, 8, and 10 were taken from an existing longitudinal database. We analyzed these for breadth by counting the number of words defined correctly and for depth by determining the amount of information in each correct definition. Via a linear mixed model, we determined whether breadth and depth varied with language diagnosis independent of nonverbal IQ, mothers' education level, race, gender, income and (for depth only) word.

Results—Children with LI scored significantly lower than children with ND on breadth and depth of vocabulary knowledge in all grades. The extent of the deficit did not vary significantly across grades. Language diagnosis was an independent predictor of breadth and depth and as strong a predictor as maternal education. For the LI group, growth in depth relative to breadth was slower than for the ND group.

Conclusions—Compared to their grade-mates, children with LI have fewer words in their vocabularies and they have shallower knowledge of the words that are in their vocabularies. This deficit persists over developmental time.

Keywords

vocabulary; developmental language impairment; definitions

Background

Vocabulary deficits are a common manifestation of developmental language impairment (LI). Children who come to be diagnosed with LI are often described as late to acquire first words (Trauner, Wulfeck, Tallal, & Hesselink, 1995) and word learning continues to be problematic for some affected preschoolers (Gray, 2004; Rice, Buhr, & Nemeth, 1990; Windfuhr, Faragher, & Conti-Ramsden, 2002), school children (Nash & Donaldson, 2005; Riches, Tomasello, & Conti-Ramsden, 2005), and adults (McGregor, Licandro, Arenas, Eden, Stiles et al., 2012). Given that the lexicon is an accumulated store of learned words, it is not surprising that the deficit is not merely one of learning words in the moment but also one of limited vocabulary knowledge (see McGregor 2009 for a review).

A useful way to think about limited vocabulary knowledge is to consider both breadth and depth. Breadth refers to how many words you know; depth to how well you know them. Vocabulary of great breadth allows precise, nuanced communication. Words known in great depth can be used flexibly in multiple contexts (Stahl, 1998); understood without contextual support (Vygotsky, 1962); and defined in detail (Dockrell & Messer, 2004). Depth also entails knowing the relationship of a given word to other words. For example, on word association tasks, paradigmatic answers, which reveal knowledge of category relationships (e.g., *insect-bee*), are taken as an indicator of depth (Schoonen & Verhallen 2008). The distinction between breadth and depth is supported by their differential value as predictors of reading achievement. Measures of breadth predict concurrent levels of decoding and word recognition whereas measures of depth predict concurrent levels of reading comprehension (Ouellette, 2006).

Deficits in depth are well-attested among individuals with LI. Compared to unaffected age-mates, they tend to provide fewer paradigmatic answers on word association tasks (Sheng & McGregor, 2010; McGregor, Berns, Owen, Michels, Duff, et al., 2011) and less information in word definitions (Marinellie & Johnson, 2002; McGregor et al., 2011). Five-to-seven year olds with LI name less well than age-mates and their naming errors are associated with sparse definitions and drawings (McGregor, Newman, Reilly, & Capone, 2002). Eleven-year olds with LI have difficulty selecting synonyms for given word targets, performing more poorly than children with normal language development who are two years their junior (Botting & Adams, 2005).

It is more problematic to determine whether a deficit in breadth is characteristic of LI. Certainly being late to talk constitutes the ultimate limitation in breadth and even after late talkers begin to use words, breadth is an obvious problem. For example, of 30 24-to-31 month-old talkers diagnosed with expressive language impairments via the Reynell Expressive Language Scale (Reynell, 1977), only 1 had a parent-reported vocabulary size of greater than 50 words (Rescorla & Ratner, 1996). However, during the preschool and early school years, vocabulary breadth is often described as a relative strength (e.g., Tomblin, McGregor, & Bean, 2011).

By far the most common way to estimate vocabulary breadth during the preschool and school years is to administer a forced-choice picture recognition task such as that used in the Peabody Picture Vocabulary Test (PPVT, Dunn & Dunn, 2007) or the British Picture Vocabulary Scale (BPVS, Dunn, Dunn, Whetton, & Pintilie, 1982). Individuals with LI often perform in the average range on such measures (Gray, Plante, Vance, & Henrichsen, 1999). For example, in McGregor et al. (2011) only 1 of 14 school children identified as having LI on the basis of scores on tests of syntax also scored more poorly than one standard deviation below the mean on the PPVT.

It could be that deficits in breadth remain in older children but our ability to identify them is undermined by reliance on forced-choice recognition measures. Leonard (2009) questioned the utility of testing comprehension in children with LI via recognition formats and Gray and colleagues (1999) recommended against the use of four specific single word vocabulary tests for diagnosis of LI, two of which employed recognition formats. Such measures are not adequately sensitive.

Alternatively, it could be that problems with breadth actually resolve over developmental time as the child who is late to talk begins to increase the rate at which he or she adds new words to the lexicon. The capacity of short term memory undergoes a two-to-three-fold expansion between the ages of 4 and 14 years and short term memory for phonological information has long been considered essential for adequate word learning (Gathercole, 1999, but see Melby-Lervåg et al., 2012). If this growth in memory capacity also characterizes development among children with LI, then some narrowing of the gap between them and their unaffected peers could be possible. Indeed, Rice (2004) reported that, for children with LI, the rate of growth in vocabulary size as measured by PPVT raw scores between ages 5 and 8 years kept pace with that of younger unaffected children. However, other evidence suggests that the gap may widen. Specifically, children identified with persistent LI at age five demonstrated declining standardized scores on the BPVS between ages 5 and 16 years (Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998).

Aims

In the current study, we aimed to fill two gaps in our understanding of vocabulary deficits associated with LI. First, we wished to determine whether the deficit involves problems with both breadth and depth. We hypothesized that both are problematic and predicted that, with a sensitive measure, both would be detected. Second, we wished to determine whether the extent of the deficit changes over developmental time. Here we did not make a firm prediction because patterns of improvement and decline have both been reported previously. To address these aims, we mined standardized test data collected as part of the Child Language Research Center Project, a longitudinal epidemiologic study of developmental LI (Tomblin, Records, Buckwalter, Zhang, Smith, & O'Brien, 1997). Specifically, we analyzed the oral definitions collected from the participants in the Center Project when they were in grades 2, 4, 8, and 10. To address breadth, we determined the number of words defined correctly, and to address depth, we determined the amount of information in each correct definition.

Methods

Participants

The 502 children who completed the expressive subtest of the *Comprehensive Receptive and Expressive Vocabulary Test* (CREVT, Wallace & Hammill, 1994) in each of grades 2, 4, 8, and 10 were selected from the database of the Child Language Research Center Project (Tomblin et al., 1997).

All of the children were monolingual American English speakers. All had performance IQ scores on the Wechsler Intelligence Scale for Children-III (WISC-III, Wechsler, 1991) of 70 or higher and normal hearing acuity as determined by a routine audiometric screening. None had emotional disturbances, frank motor impairments, or neurological deficits per parent and teacher reports. Demographic characteristics of the sample as determined via parent report appear in table 1.

We classified the children according to their kindergarten language diagnoses. In kindergarten, the children completed a test battery that yielded five composite scores: receptive language, expressive language, vocabulary, grammar, and narrative. The CREVT was not administered in the kindergarten year; therefore, the children's diagnoses were not confounded with performance on the CREVT definitions. A child met criteria for LI if two or more of these composite scores were −1.25 or poorer given local norms established through the Child Language Research Center Project (Tomblin, Records, & Zhang, 1996). Of the LI group, 28% demonstrated expressive deficits only; 19% demonstrated receptive deficits only; and 53% demonstrated deficits in both modalities (Tomblin, Zhang, Buckwalter, & O'Brien, 2003).

It should be noted that some children changed LI or ND classifications over the years of the project (Tomblin et al., 2003). Of the 177 children classified as LI in kindergarten, 116

(66%) were classified as LI in two or more subsequent grades. Of the 325 children classified as ND in kindergarten, 307 (95%) were classified as ND in two or more subsequent grades. That said, classification according to kindergarten diagnosis alone is the correct approach given that we were interested in determining potential improvements over developmental time.

We did not subgroup the children with LI according to whether they had specific LI (SLI) or nonspecific LI (NLI), the typical distinction being that those with SLI have nonverbal IQs of at least 85 whereas those with NLI have nonverbal IQs of 70–84. Previous analysis of the data in the Child Language Research Center Database revealed similar language profiles for the SLI and NLI subgroups (Tomblin, Zhang, Catts, Ellis Weismer, & Weiss, 2004). Moreover, the response of children with SLI and NLI to oral language interventions that include a vocabulary focus do not differ (Bowyer-Crane, Duff, Hulme, & Snowling, 2011). However, to determine whether the effect of language diagnosis on vocabulary knowledge was independent of nonverbal IQ, we did enter two levels of nonverbal IQ as a predictor variable, higher being average or above (85+) and lower being below average (70-to-84). Seventy-seven of 177 children (44%) in the LI group had below average nonverbal IQs as did 78 of 325 children (24%) in the ND group.

Materials

The CREVT is a standardized measure of oral vocabulary for children ages 4;0 to 17;11 normed on 1,920 children from 32 of the 50 United States. The expressive subtest requires children to define up to 25 words. The words are all nouns although four of them (*shampoo, roar, chisel, perfume*) can be used as verbs as well. With the exception of *roar*, all of the nouns label people or concrete objects. Children younger than 12 years begin with the first item (*knife*) and continue until a ceiling of three consecutive incorrect responses is reached or until they reach the final item (*teamster*). Children who are 12 and older begin with item 14 (*kettle*) and also continue until they reach a 3-item ceiling or until they reach the final item. There is also a basal such that children who do not respond with three consecutive correct responses drop to lower numbered items until they do. In the standardized procedure, the child receives 1 point for each correct definition and 0 for incorrect definitions. The number of correct responses constitutes the raw score.

Procedure

The children's responses on the CREVT were transcribed from the audio-recordings of the test session and the hand-written test forms. These responses were then scored with a scheme first reported in McGregor et al. (2011) that was designed to capture subtle differences in depth of word knowledge. Definitions received a 1 if they included words that bore some meaningful relationship to the target but did not define it with a minimal level of precision. Conventional but minimal definitions received a 2 and definitions that included more than minimal correct information received 3. If words had multiple meanings, either was acceptable. For example, the following definitions of *mandolin* are arranged by score from lowest to highest: "an instrument" (1); "a musical instrument with strings" (2); "an instrument used to finely slice or grate, it's a kitchen tool that has sharp blades" (3).

Each definition was scored by three independent scorers. To calibrate, they first determined scoring rubrics by consulting dictionaries and thesauruses. They determined, for example, that to receive a 3, the definition of the musical instrument *mandolin* had to include *instrument* and *strings* and either *pear-shaped* or *plucked* or, if the child was aiming for the kitchen version, "*kitchen utensil* and either *slicing* or *grating*." Scores that differed amongst the three scorers by a single point were averaged. For example, if one definition was scored as 0, 0, and 1, the assigned score was .33. Therefore, the potential range of depth scores was .33 to 3. Scores that differed by more than a single point were discussed so that a consensus score could be determined. This happened in 291 of 25,681 definitions for an overall level of agreement +/−1 of 98.87%.

Data analysis

We conducted three analyses. In the first, the dependent variable was the number of words defined correctly. In accord with the basal scoring system of the CREVT, we counted any definitions below basal as correct as well. Thus this dependent variable was equivalent to the raw score on the CREVT. Given the binary nature of this value—correct or incorrect—this dependent variable was taken as an indication of breadth. In the second analysis, the dependent variable was the value determined by the three independent scorers for each correct definition produced. Because the scorers used a continuous scale to capture the amount of information, this dependent variable was taken as an indication of depth. The third analysis employed the same dependent variable as the second. What differed was that we included the data from only those children who had earned a score of 3 on at least one of their definitions in second grade. We did this to determine whether we still obtained effects of language diagnosis on the depth scores if we limited the analysis to those children who could formulate a good definition.

In all three analyses, the primary goal was to determine whether scores varied with language diagnosis. We also determined other independent predictors. Fixed predictor variables that did not change over time were diagnosis (2 levels), nonverbal IQ (2 levels), maternal education level (3 levels), race (2 levels), gender (2 levels), and income (7 levels treated as continuous). Grade (4 levels), of course, changed over time. An interaction term between diagnosis and grade level was included to determine whether the effect of diagnosis changed over developmental time. An interaction term between nonverbal IQ and grade level was also included to determine whether the effect of nonverbal IQ changed over developmental time. In the second and third analyses, word was also a predictor and, because of basals and ceilings, word changed over time. For all three analyses, we used a linear mixed model (LMM) to analyze the longitudinal dataset. Analyses were carried out in Proc Mixed of SAS v9.3.

Results

The ranges of breadth and depth scores appear in Table 2. The CREVT appeared adequately sensitive to potential group differences and growth in breadth as the percentage of children scoring at floor and ceiling levels in any group at any given grade was typically zero and never greater than 4%. As for depth, fewer than 2% of children in any group at any given grade scored at floor. Sensitivity at the upper end was reduced somewhat in comparison with 9 – 15% of participants scoring at ceiling per group and grade.

Breadth

We first analyzed the breadth score, the number of words correctly defined. Scores ranged from 3 to 25 and each child had scores at 4 different grade levels. An unstructured correlation matrix was used to model within subject correlation over time which allowed each grade combination to have a different correlation. The resulting correlation matrix revealed that the breadth scores in second grade had a correlation of 0.437 with breadth scores in fourth grade, 0.353 with eighth grade, and 0.331 with tenth grade. Breadth scores in fourth grade had a correlation of 0.470 with breadth scores in eighth grade and 0.412 in tenth grade. Finally, the eighth to tenth grade correlation was 0.638.

Average breadth score was significantly influenced by language diagnosis, grade level, nonverbal IQ, gender, race, maternal education level, and income (see table 3). Children with normal language development scored higher than those with LI at all grades and children at each grade scored higher than children at the preceding grade. From grade 2 to 4, the scores increased on average 33% ($p<0.0001$). From grade 4 to 8, there was an increase of 25% ($p<0.0001$). The change from grade 8 to 10 was a 0.06% increase ($p<0.0001$). There was no significant grade x diagnosis interaction suggesting that differences between the diagnostic groups remained consistent across grade levels. From grades 2 to 10, the average breadth scores for the LI group grew from 9.32 (SD = 3.25) to 17.54 (SD = 2.88); for the ND group they grew from 12.28 (3.44) to 20.73 (2.68). On average, children with normal language development scored 2.25 words higher than children with LI in second grade, 2.09 in fourth grade, 2.14 in eighth grade, and 2.42 in 10^{th} grade (see figure 1).

As for other predictors, children with higher nonverbal IQs scored 1.00 words higher than those with lower nonverbal IQs and there was no grade x nonverbal IQ interaction. Girls scored 0.573 words higher than boys and children of majority race scored 0.954 words higher than children of minority races. Children whose mothers had 17 or more years of education scored 1.074 words higher ($p=0.0044$) than children whose mothers had 13 to 16 years and the children whose mothers had 13 to 16 years scored 0.626 words higher than children who had less than 13 years ($p=0.0032$). Increased income also led to an increase in score.

Depth

Next we analyzed depth scores. Recall that, for each word defined correctly, the depth score was the average rating assigned by three scorers who used a scale of 1 to 3 to quantify amount of information. The minimum number of words defined correctly was 3 and the maximum was 25. Thus, depth score was repeated between 3 and 25 times at each grade with each child having scores at 4 different grade levels. Word was treated as a random effect with compound symmetry. The correlation structure was more complicated in this analysis because each subject had correlation between grades and also correlation between words within grade. Compound symmetry was used for both structures which assumed all pairwise comparisons had the same correlation.

As seen in Table 4, depth varied with diagnosis such that children with normal language had higher depth scores than children with LI. There was a significant increase in scores per correctly defined word over grades. From grade 2 to 4, the scores increased on average 12% ($p<0.0001$). From grade 4 to 8, there was an increase of 10% ($p<0.0001$). The change from grade 8 to 10 was a 3% increase ($p=0.0008$). The grade x diagnosis interaction was not significant. Children with normal language development scored 0.14, 0.16, 0.16, 0.13 higher than children with LI at grades 2, 4, 8, 10 respectively (see figure 2).

Average depth score was also significantly influenced by nonverbal IQ, gender, maternal education level, income, and word. Children of majority race did not score significantly higher than children of minority races. On average and per word, children with higher nonverbal IQs scored 0.052 higher than those with lower nonverbal IQs and there was no grade x nonverbal IQ interaction. Girls scored 0.035 higher than boys. Children whose mothers had more than 16 years of education scored 0.095 points higher ($p = 0.0005$) than children whose mothers had 13 to 16 years and children whose mothers had 13 to 16 years scored 0.063 points higher ($p < 0.0001$) than children whose mothers had less than 13 years. Increased income led to an increase in score. In general, words later in the order of administration elicited lower scores than those that appeared earlier. Of those words that scored more than zero, the easiest word was word 7 (*envelope*) while the most challenging word was 22 (*mandolin*).

Depth determined only for children who could formulate good definitions

The results of the second analysis suggest that children with LI have shallower word knowledge than their unaffected grade-mates. However, it is possible that the children with LI knew just as much about the words as their peers but they were less able to formulate definitions. To explore this possibility, we asked whether the diagnostic groups still differed if we included only the children who showed direct evidence of being able to formulate a good definition, that is, only the children who had earned a score of 3 on one of their definitions at grade 2. Meeting this criterion were 97 children with LI and 224 children with normal language development. The analysis was otherwise identical to the original analysis of depth scores.

As seen in Table 5, depth still varied with diagnosis such that children with normal language development had larger depth scores than children with LI. There was a significant increase in scores per correctly defined word over grades. From grade 2 to 4, the scores increased on average 6% ($p<0.0001$). From grade 4 to 8, there was an increase of 8% ($p<0.0001$). The change from grade 8 to 10 was a 4% increase ($p=0.0011$). Again, there was no significant grade x language diagnosis interaction. Children with normal language development scored 0.11, 0.16, 0.17, 0.16 higher than children with LI at grades 2, 4, 8, 10 respectively (see Figure 3).

As before, average depth score was also significantly influenced by nonverbal IQ, gender, maternal education level, income, and word. Children of majority race were not significantly different from children of minority races. On average and per word, children with higher nonverbal IQ scored 0.05 higher than those with lower nonverbal IQ; there was no grade x nonverbal IQ interaction. Girls scored 0.053 higher than boys. Children whose mothers had more than 16 years of education scored 0.090 points higher than children whose mothers had 13 to 16 years ($p = 0.0024$) and children whose mothers had 13 to 16 years scored 0.067 points higher ($p = 0.0002$) than children whose mothers had less than 13 years. Increased income led to an increase in score. In general, words later in the order of administration elicited lower scores than those that appeared earlier. The easiest word was word 7 (*envelope*) while the most challenging word was 24 (*macaw*).

The relationship between breadth and depth

Finally, we examined individual variation within the groups to determine whether the children who had the greatest strengths (or weaknesses) in breadth were the same children who had the greatest strengths (or weaknesses) in depth. Because the depth scores are not balanced across words and grades, a simple correlation between breadth and depth is not appropriate. We constructed a linear mixed model to predict depth scores from the breadth scores. In this way, we could account for the imbalance in number of words scored per child to determine the relationship between breadth and depth. Similar to our previous analysis, the LMM controlled for grade level and diagnosis category and included an interaction between diagnosis and breadth score so that the relationship between breadth and depth could differ between the groups. All variables in the model were statistically significant ($p<.0001$). Both the breadth variable ($b=.078$, $p<.0001$) and the interaction term ($b=.0156$, $p<.0001$) were significant indicating that those children who had higher breadth scores also had higher depth scores and that the relationship between breadth score and depth score varied between the ND and LI groups. Specifically, for every one word increase in breadth score in LI children, the depth score increased by 0.078, on average. For ND children, every one word increase in breadth score related to depth score increasing by 0.0938, on average.

Discussion

This analysis of definitions collected longitudinally at grades 2, 4, 8, and 10 from 177 children with LI and 325 unaffected grade-mates confirmed the hypothesis that children with LI have vocabulary deficits characterized by limited breadth and depth of word knowledge. These deficits persist throughout the school years.

Vocabulary development is highly variable across individuals. Here we demonstrated that vocabulary knowledge varies with grade level as is logical given accumulating exposure to words over time. We also demonstrated moderate effects of maternal education and the related variable income, as consistent with a large literature linking maternal education and other socioeconomic variables with the amount and type of vocabulary input children receive (Dollaghan, Campbell, Paradise, Feldman, Janosky, et al., 1999; Hoff, 2003). However, maternal education is not a predictor of all aspects of language development. Using the same database employed here, Rice and colleagues reported no link between maternal education level and children's performance on a grammatical tense marking task (Rice, Tomblin, Hoffman, Richman, & Marquis, 2004). The specificity of the relationship between levels of maternal education and vocabulary development reflects the highly experiential nature of word learning.

To a smaller extent, vocabulary knowledge was affected by nonverbal IQ, gender, and race. Although the range of nonverbal IQs was limited to 70 and above, those with average or higher IQs demonstrated greater breadth and depth of word knowledge than those with below average IQs. This finding is consistent with previously reported associations between nonverbal IQ and performance on composite measures of language ability (DeThorne & Watkins, 2006) and vocabulary (Hammill, Pearson, Wiederholt, 1997). Specifically, the association prompts hypotheses about the role of nonverbal cognitive processes in word learning. Features of the visual context (e.g., where a speaker is looking while using a word) and features of word referents themselves (e.g., the shape of a given object) are known to support word learning (Hollich, Hirsh-Pasek, & Golinkoff, 2000) and there may be important individual differences in the processing of such information. It could also be that domain general mechanisms, such as the ability to extract statistical patterns, are useful in both verbal and nonverbal tasks (Kirkham, Slemmer, & Johnson, 2002).

Girls outperformed boys on both breadth and depth. It is unusual to see reports of gender differences in vocabulary knowledge during the school years. Although girls typically have larger vocabularies than boys before age two years, these differences are reported to disappear shortly thereafter (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991). Norm-referenced measures of vocabulary reflect this state of affairs. Consider, for example, that the MacArthur-Bates Communicative Development Inventory (Fenson, Dale, Reznick, Thal, Bates, Hartung, et al., 1991) designed for children of 30 months and younger, has separate norms for boys and girls whereas the PPVT, designed for people ages 30 months to 90+ years, does not. It could be that the difference between girls and boys reflects the girls' better facility with definitions than stronger vocabulary knowledge *per se* but recall that the gender effect remained when poorer definers were removed from the sample. Instead, because the gender effect is small, it has likely been missed in previous studies that lack sensitive measures or a large sample size.

Finally, word knowledge varied with race with majority outperforming minority but this was characteristic of breadth only. This likely reflects differences between communities in the words used, differences that are often overlooked in test construction (Campbell, Dollaghan, Needleman, & Janowsky, 1999). Because depth was calculated only for known words, race was no longer predictive.

With these other factors as a backdrop, we can better interpret the extent to which vocabulary development varies with LI. In terms of breadth and depth, the effect of language diagnosis was moderate. In fact, diagnosis exerted as large an influence as maternal education. A diagnosis of LI is an important independent predictor of vocabulary breadth and depth.

In terms of breadth, children with LI knew fewer words than their unaffected grade-mates at all grades. This must be interpreted in light of the task, an expressive task that required explanation of the word meaning. To be conservative, we can say that, as a group, school children with LI know fewer words than their peers when “knowing” is operationalized as being able to say at least one correct bit of information about what that word means.

Given this deficit and its manifestation at all grade levels, expanding vocabulary breadth by introducing new words will be an important clinical goal for many children with LI throughout schooling. This goal is motivated by the links between vocabulary size and popularity with peers (Gertner, Rice, & Hadley, 1994), success in reading (Clarke, Snowling, Truelove, & Hulme, 2010), and achievement in academic tasks (Dollinger, Matyja, & Huber, 2008). Potential new words can be selected for their functionality, relevance to academic tasks, developmental appropriateness, and teachability and they can be taught via incidental exposure, direct teaching, and coaching in the use of context to infer new word meanings (McGregor & Duff, in press).

As determined by ratings of their definitions, children with LI demonstrated problems with depth at all grades as well. Of course the ability to provide a good definition depends on metalinguistic and syntactic abilities as well as knowledge of word meaning. There is reason to worry about these confounds. For example, Marinellie and Johnson (2002) found third-, fourth-, and fifth-graders with LI to have difficulty not only with the content of definitions but also with their form. It could be that problems with form limit the child’s ability to express perfectly adequate knowledge of word meaning. We addressed the potential confound between word knowledge and ability to formulate definitions by re-running the depth analysis after excluding all participants who were unable to demonstrate a single highly-rated definition in second grade. The difference in depth scores between children with and without LI remained; in fact, the effect was larger than before.

We conclude that, as a group, children with LI have shallower knowledge of word meanings than their unaffected grade-mates throughout the school years. This finding is consistent with previous findings based on other children with LI in our own laboratory (McGregor et al., 2002; McGregor et al., 2011) and in an independent laboratory (Marinellie & Johnson, 2002). Children with LI may need support for enriching their word knowledge. In many communicative and academic settings, it is not enough to be merely familiar with words, rather, a deep appreciation is necessary. Consider the difficulty in understanding humor, winning an argument, or understanding a newspaper article without a good grasp of the meanings of the words involved. Knowledge of familiar words can be deepened via activities that require comparison and contrast of word meanings, generation of synonyms and antonyms, and use of target words in a variety of contexts. Visuals such as word walls, word maps, and Venn diagrams can be useful supports for these activities (McGregor & Duff, in press).

For the children in the LI and ND groups, breadth and depth were associated such that children with higher breadth scores also had higher depth scores. This coupling has clinical implications in that the child who needs intervention focused on depth will likely need intervention focused on breadth as well. This is to be expected as the constructs of breadth and depth are not fully independent. Imagine, for example, the impossibility of fully knowing what *dog* means without knowing the meaning of *animal*, *pet*, *bark,* and *cat*. The more words you know, the more you can appreciate what a word means in relation to other words and this inter-relatedness serves to deepen knowledge for any given word (Meara & Wolter, 2004; Vermeer, 2001). Although breadth and depth were associated in both groups, growth in depth relative to breadth was slower for the LI group than the ND group. This finding too holds clinical implications. Clinicians should pay attention to both dimensions of vocabulary development but they may find children with LI to require even more support for deepening than for broadening their vocabulary knowledge.

Having established that children with LI are deficient relative to peers with ND in breadth and depth of word knowledge throughout the school years, we turn to the question of stability. Hypothetically, the gap between LI and ND groups could narrow or widen over developmental time. Neither change was evident in the present study. However, before accepting the null, the limitations of the CREVT as a tool for detecting relative change must be considered. Floor and ceiling level performances could limit detection yet this was clearly not a problem for breadth scores and only a minor problem for depth scores given that at least 85% of the children remained below ceiling in depth scores at all grades. A more likely limitation concerns the lack of verbs and abstract nouns on the CREVT. Such items might better differentiate LI and ND groups, especially in the later school years and, therefore, might be more sensitive to a widening of the gap between them over time. This hypothesis should be tested in future studies. In the present study, deficits in breadth and depth relative to unaffected peers were stable throughout the school years.

Conclusions

As a group, children with LI present with deficient vocabulary knowledge characterized by limitations in both breadth and depth. The association between LI and deficient vocabulary knowledge is independent of maternal education, income, nonverbal IQ, gender, and race, but vocabulary also varies in relation to these factors.

The vocabulary deficit that characterizes children with LI in second grade persists over the school years. Given hundreds of thousands of words in any given language, vocabulary development is an open-ended, life-long enterprise. The word learner with LI may be unlikely ever to close the gap fully.

Acknowledgments

We gratefully acknowledge J. Bruce Tomblin's generosity in sharing his data and providing comments on the initial draft of this paper. Melissa Lloyd provided assistance with data coding. The first author was supported by awards from the National Institutes of Health during data analysis (2R01DC003698) and manuscript preparation (1R01DC011742).

References

BOTTING N, ADAMS CV. Semantic and inferencing abilities in children with communication disorders. International Journal of Language and Communication Disorders. 2005; 40:49–66. [PubMed: 15832525]

BOWYER-CRANE C, DUFF F, HULME C, SNOWLING MJ. The response to intervention of children with SLI and general delay. Learning Disabilities: A Contemporary Journal. 2011; 9:5–19.

CAMPBELL T, DOLLAGHAN C, NEEDLEMAN H, JANOSKY J. Reducing bias in language assessment: Processing dependent measures. Journal of Speech, Language, and Hearing Research. 1997; 40:519–525.

CLARKE PJ, SNOWLING MJ, TRUELOVE E, HULME C. Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. Psychological Science. 2010; 21:1106–1116. [PubMed: 20585051]

DETHORNE LS, WATKINS R. Language abilities and nonverbal IQ in children with language impairment: Inconsistency across measures. Clinical Linguistics & Phonetics. 2006; 20:641–658. [PubMed: 17342874]

DOCKRELL, JE.; MESSER, D. Lexical acquisition in the early school years. In: Berman, RA., editor. Language development across childhood and adolescence. Philadelphia: John Benjamins; 2004. p. 35-52.

DOLLAGHAN CA, CAMPBELL TF, PARADISE JL, FELDMAN HM, JANOSKY JE, PITCAIRN DN, et al. Maternal education and measures of early speech and language. Journal of Speech, Language, and Hearing Research. 1999; 42:1432–1443.

DOLLINGER SJ, MATYJA AM, HUBER JL. Which factors best account for academic success: Those which college students can control or those they cannot? Journal of Research in Personality. 2008; 42:872–885.

DUNN, LM.; DUNN, LM.; WHETTON, C.; PINTILIE, D. British Picture Vocabulary Scale. Windsor, England: NFER-Nelson; 1982.

DUNN, M.; DUNN, DM. Peabody Picture Vocabulary Test. 4. Bloomington, MN: Pearson; 2007.

FENSON, L.; DALE, PS.; REZNICK, JS.; THAL, D.; BATES, E.; HARTUNG, JP., et al. Technical manual for the MacArthur Communicative Development Inventories. San Diego, CA: San Diego State University Press; 1991.

GATHERCOLE SE. Cognitive approaches to the development of short-term memory. Trends in Cognitive Sciences. 1999; 3:410–419. [PubMed: 10529796]

GERTNER BL, RICE ML, HADLEY PA. Influence of communicative competence on peer preferences in a preschool classroom. Journal of Speech and Hearing Research. 1994; 37:913–923. [PubMed: 7967575]

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

GRAY S, PLANTE E, VANCE R, HENRICHSEN M. The diagnostic accuracy of four vocabulary tests administered to preschool-age children. Language, Speech, and Hearing Services in Schools. 1999; 30:196–206.

HAMMILL, DD.; PEARSON, NA.; WIEDERHOLT, JL. Examiners manual: Comprehensive Test of Nonverbal Intelligence. Austin, TX: pro-ed; 1997.

HOFF E. The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. Child Development. 2003; 74:1368–1378. [PubMed: 14552403]

HOLLICH GJ, HIRSH-PASEK K, GOLINKOFF RM. Breaking the language barrier: An emergentist coalition model for the origins of word learning. Monographs of the Society for Research in Child Development. 2000; 65(3) Serial No. 262.

HUTTENLOCHER J, HAIGHT W, BRYK A, SELTZER M, LYONS T. Early vocabulary growth: Relation to language input and gender. Developmental Psychology. 1991; 27:236–248.

KIRKHAM NZ, SLEMMER JA, JOHNSON SP. Visual statistical learning in infancy: Evidence for a domain general learning mechanism. Cognition. 2002; 83:B35–B42. [PubMed: 11869728]

LEONARD LB. Is expressive language disorder an accurate diagnostic category? American Journal of Speech-Language Pathology. 2009; 18:115–123. [PubMed: 19029534]

MARINELLIE SA, JOHNSON CJ. Definitional skill in school-age children with specific language impairment. Journal of Communication Disorders. 2002; 35:241–259. [PubMed: 12064786]

MCGREGOR, KK. Semantic deficits associated with developmental language disorders. In: Schwartz, RG., editor. The Handbook of Child Language Disorders. New York: Psychology Press; 2009. p. 365-387.

MCGREGOR KK, BERNS A, OWEN AJ, MICHELS S, DUFF D, BAHNSEN AJ, LLOYD M. Associations between syntax and the lexicon among children with or without ASD and language impairment. Journal of Autism and Developmental Disorder. 2011; 42:35–47.

MCGREGOR, KK.; DUFF, D. Promoting diverse and deep vocabulary development. In: Ukrainetz, T., editor. School speech-language pathology: Principles and procedures for oral and written language. Austin, TX: Pro-Ed; in pressTo appear

MCGREGOR, KK.; LICANDRO, U.; ARENAS, R.; EDEN, N.; STILES, D.; BEAN, A.; WALKER, E. Why Words are Hard for Adults with Developmental Language Impairments. 2012. Manuscript under review

MCGREGOR KK, NEWMAN RM, REILLY RM, CAPONE N. Semantic representation and naming in children with specific language impairment. Journal of Speech, Language, and Hearing Research. 2002; 45:998–1014.

MEARA P, WOLTER B. V-links: beyond vocabulary breadth. Angles on the English Speaking World. 2004; 4:85–96.

MELBY-LERVÅG M, LERVÅG A, LYSTER SAH, KLEM M, HAGTVET B, HULME C. Nonword-Repetition ability does not appear to be a causal influence on children's vocabulary development. Psychological Science. 2012; 23:1092–1098. [PubMed: 22923338]

NASH M, DONALDSON ML. Word learning in children with vocabulary deficits. Journal of Speech, Language, and Hearing Research. 2005; 48:439–458.

OUELLETTE GP. What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. Journal of Educational Psychology. 2006; 98:554–566.

RESCORLA L, RATNER NB. Phonetic profiles of toddlers with specific expressive language impairment. Journal of Speech and Hearing Research. 1996; 39:153–165. [PubMed: 8820707]

REYNELL, J. Reynell Developmental Language Scales. Windsor: NFER; 1977.

RICE, M. Growth models of developmental language disorders. In: Rice, ML.; Warren, SF., editors. Developmental language disorders: From phenotypes to etiologies. Mahwah, NJ: Erlbaum; 2004. p. 207-240.

RICE ML, BUHR J, NEMETH M. Fast mapping word-learning abilities of language delayed preschoolers. Journal of Speech and Hearing Disorders. 1990; 55:33–42. [PubMed: 2299838]

RICE ML, TOMBLIN JB, HOFFMAN L, RICHMAN WA, MARQUIS J. Grammatical tense deficits in children with SLI and nonspecific language impairment: Relationships with nonverbal IQ over time. Journal of Speech, Language, and Hearing Research. 2004; 47:816–834.

RICHES NG, TOMASELLO M, CONTI-RAMSDEN G. Verb learning in children with SLI: frequency and spacing effects. Journal of Speech, Language, and Hearing Research. 2005; 48:1397–1411.

SCHOONEN R, VERHALLEN M. The assessment of deep word knowledge in young first and second language learners. Language Testing. 2008; 25:211–236.

SHENG L, MCGREGOR KK. Lexical-semantic organization in children with specific language impairment. Journal of Speech, Language, and Hearing Research. 2010; 53:146–159.

STAHL, SA. Four questions about vocabulary knowledge and reading and some answers. In: Hynd, CR.; Stahl, SA.; Carr, M.; Glynn, SM., editors. Learning from text across conceptual domains. Mahwah, NJ: Lawrence Erlbaum Associates, Inc; 1998. p. 73-94.

STOTHARD SE, SNOWLING MJ, BISHOP DVM, CHIPCHASE BB, KAPLAN CA. Language impaired preschoolers: A follow-up into adolescence. Journal of Speech, Language, and Hearing Research. 1998; 41:407–418.

TOMBLIN, JB.; MCGREGOR, KK.; BEAN, A. Specific language impairment. In: Amaral, DG.; Dawson, GD.; Geschwind, DH., editors. Autism Spectrum Disorders. Oxford University Press; 2011. p. 315-329.

TOMBLIN JB, RECORDS NL, BUCKWALTER P, ZHANG X, SMITH E, O'BRIEN M. Prevalence of specific language impairment in kindergarten children. Journal of Speech and Hearing Research. 1997; 40:1245–1260.

TOMBLIN JB, RECORDS NL, ZHANG X. A system for the diagnosis of specific language impairment in kindergarten children. Journal of Speech and Hearing Research. 1996; 39:1284–1294. [PubMed: 8959613]

TOMBLIN JB, ZHANG X, BUCKWATER P, O'BRIEN M. The stability of primary language disorder: Four years after kindergarten diagnosis. Journal of Speech, Language, and Hearing Research. 2003; 46:1283–1296.

TOMBLIN, JB.; ZHANG, X.; CATTS, H.; ELLIS WEISMER, S.; WEISS, A. Dimensions of individual differences in communication skills among primary grade children. In: Rice, ML.; Warren, SF., editors. Developmental language disorders: From phenotypes to etiologies. Mahwah, NJ: Erlbaum; 2004. p. 53-76.

TRAUNER, D.; WULFECK, B.; TALLAL, P.; HESSELINK, J. Technical Report, Publication No CND-9513. San Diego, CA: Center for Research in Language, UCSD; 1995. Neurologic and MRI profiles of language impaired children.

VERMEER A. Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. Applied Psycholinguistics. 2001; 22:217–234.

VYGOTSKY, L. Thought and language. Cambridge, MA: MIT Press; 1962.

WALLACE, G.; HAMMILL, D. Comprehensive Receptive and Expressive Vocabulary Test. Austin, TX: Pro-Ed; 1994.

WECHSLER, D. Wechsler Intelligence Scale for Children. 3. San Antonio, TX: The Psychological Corporation; 1991.

WINDFUHR KL, FARAGHER B, CONTI-RAMSDEN G. Lexical learning skills in young children with specific language impairment (SLI). International Journal of Language and Communication Disorders. 2002; 37:415–432. [PubMed: 12396842]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

What this paper adds

What is already known on this subject

Vocabulary deficits are a common manifestation of developmental language impairment (LI). Deficits in depth of word knowledge are well-attested but documentation of deficits in breadth has been more elusive. Moreover, there are conflicting reports concerning the developmental course of the vocabulary deficit. There are some indications that the deficit resolves but other evidence points to increasingly severe deficits relative to peers over time.

What this study adds

We analyzed oral definitions collected as part of a longitudinal study of children with LI. As a group, the children with LI presented with deficits in the breadth and depth of vocabulary knowledge throughout the school years. The extent of the problem relative to peers did not vary significantly over developmental time. Clinicians should be aware of this developmental pattern as it suggests that vocabulary intervention will be required throughout the school years for many children with LI.

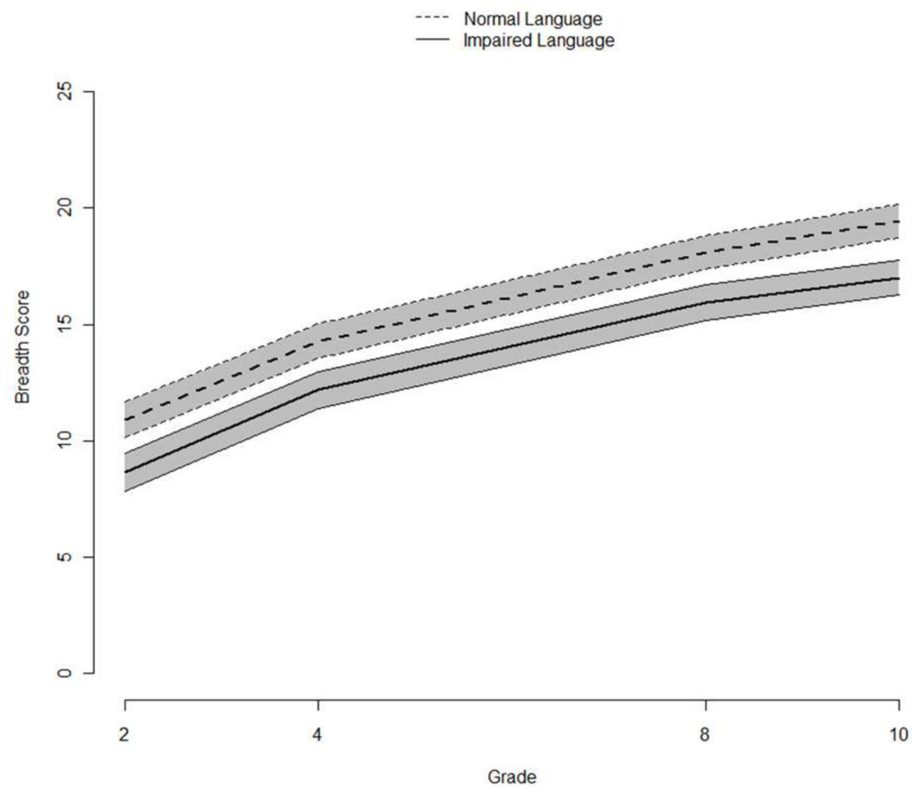

Figure 1.
Breadth score by diagnosis and grade. The shading represents the 95% confidence interval. This plot holds constant other variables such that nonverbal IQ = high, gender = male, race = majority, maternal education = 13 to 16 years, and income level = average.

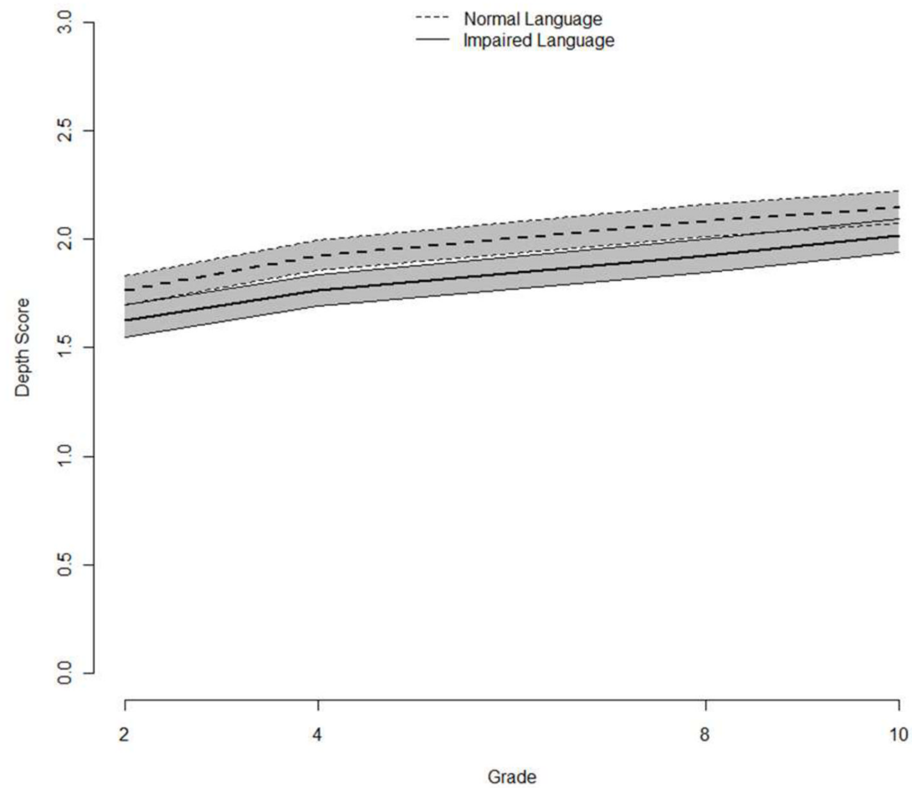

Figure 2.
Depth score by diagnosis and grade. The shading represents the 95% confidence interval. This plot holds constant other variables such that nonverbal IQ = high, gender = male, race = majority, maternal education = 13 to 16 years, income level = average, first word administered = 1, basal word = 1.

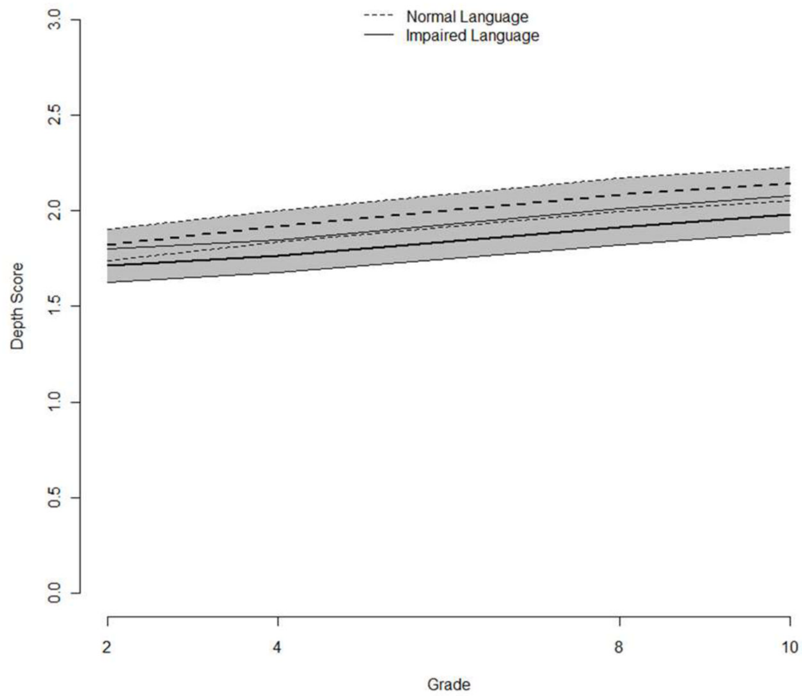

Figure 3.
Depth score by diagnosis and grade for those children who scored at least one “3” in second grade. The shading represents the 95% confidence interval. This plot holds constant other variables such that nonverbal IQ = high, gender = male, race = majority, maternal education = 13 to 16 years, income level = average, first word administered = 1, basal word = 1.

Table 1

Demographic characteristics of the sample.

DX	# Boys	# Girls	# Majority	# Minority	Years of Maternal Education M (SD)	Income* M (SD)
LI	82	70	113	39	12.62 (1,87)	2.44 (1.69)
ND	198	152	314	36	13.87 (2.17)	3.54 (1.52)

*Income was categorized on a continuous scale ranging from 1 to 7.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Range of breadth and depth scores and percentage of children scoring at floor and ceiling by grade and diagnostic group.

	Grade	LI			ND		
		range	**% at floor**	**% at ceiling**	**range**	**% at floor**	**% at ceiling**
breadth	2	3 – 18	0	0	3 – 23	0	0
depth		.33 – 3	1.27	10.42	.33 – 3	1.17	12.43
breadth	4	4 – 21	0	0	8 – 23	0	0
depth		.33 – 3	.96	10.26	.33 – 3	.98	14.49
breadth	8	12 – 23	0	0	13 – 25	0	.62
depth		.33 – 3	1.84	9.11	.33 – 3	1.27	11.25
breadth	10	12 – 24	0	0	13 – 25	0	3.75
depth		.33 – 3	1.64	9.52	.33 – 3	1.55	11.52

Note: The potential range is 0 –25 for breadth scores and .33 – 3 for depth scores.

Table 3

Test of fixed effects with number of correct definitions as the dependent variable.

Effect	Level	Mean Estimate	Effect Size	F Value	p
grade	2	10.7551	0.78	842.42	<.0001
	4	14.3306	0.82		
	8	17.8663	0.27		
	10	19.00830			
gender	M	15.2036	0.18	8.46	0.0038
	F	15.7765			
race	Majority	15.9673	0.30	10.64	0.0012
	Minority	15.0129			
diagnosis	ND	16.6034	0.35	108.46	<.0001
	LI	14.3768			
maternal education	<13	14.7151	0.23	10.31	<.0001
	13–16	15.3406			
	17+	16.4145	0.40		
income		0.2362		13.38	0.0003
nonverbal IQ	Low	14.9898	0.13	18.69	<.0001
	High	15.9904			
grade x diagnosis	2 ND	11.8824	0.54	0.76	0.5157
	2 LI	9.6277			
	4 ND	15.3780	0.55		
	4 LI	13.2833			
	8 ND	18.9338	0.59		
	8 LI	16.7988			
	10 ND	20.2193	0.70		
	10 LI	17.7973			
grade x nonverbal IQ	2 Low	10.3001	0.22	1.69	0.1687
	2 High	11.2100			
	4 Low	14.0120	0.16		
	4 High	14.6492			
	8 Low	17.2804	0.31		

Effect	Level	Mean Estimate	Effect Size	F Value	p
	8 High	18.4523			
	10 Low	18.3666	0.36		
	10 High	19.6499			

Note: The Mean Estimate column gives the estimated dependent variable value for categorical variables and the slope for continuous variables (income) computed assuming an average value of every other effect. The Effect Size (mean difference divided by a pooled standard deviation) column compares 2 levels of the effect. The F test is an overall test for the effect.

Table 4

Test of fixed effects with amount of information in each correct definition divided by total number of correct definitions as the dependent variable.

Effect	Level	Mean Estimate	Effect Size	F value	p
grade	2	1.3979	0.43	156.93	<.0001
	4	1.5598	0.41		
	8	1.7142	0.14		
	10	1.7665			
gender	M	1.5920	0.14	5.86	0.0155
	F	1.6272			
race	Majority	1.6117	0.02	0.04	0.8501
	Minority	1.6075			
diagnosis	ND	1.6841	0.30	87.04	<.0001
	LI	1.5351			
maternal education	<13	1.5358	0.28	17.44	<.0001
	13–16	1.5991			
	17+	1.6939	0.43		
income		0.0130		7.37	0.0042
word				101.45	<.0001
nonverbal IQ	Low	1.5838	0.08	9.07	0.0026
	High	1.6354			
grade x diagnosis	2 ND	1.4686	0.44	0.60	0.6147
	2 LI	1.3271			
	4 ND	1.6406	0.54		
	4 LI	1.4790			
	8 ND	1.7948	0.49		
	8 LI	1.6335			
	10 ND	1.8323	0.40		
	10 LI	1.7008			
grade x nonverbal IQ	2 Low	1.3721	0.15	1.56	0.1971
	2 High	1.4236			
	4 Low	1.5454	0.09		
	4 High	1.5742			

Effect	Level	Mean Estimate	Effect Size	F value	p
	8 Low	1.6947	0.12		
	8 High	1.7337			
	10 Low	1.7230	0.26		
	10 High	1.8101			

Note: The Mean Estimate column gives the estimated dependent variable value for categorical variables and the slope for continuous variables (income) computed assuming an average value of every other effect. The Effect Size (mean difference divided by a pooled standard deviation) column compares 2 levels of the effect. The F test is an overall test for the effect.

Table 5

Test of fixed effects with amount of information in each correct definition divided by total number of correct definitions as the dependent variable. Only participants who scored at least one "3" in second grade are included.

Effect	Level	Mean Estimate	Effect Size	F Value	Pr > F
grade	2	1.5078	0.25	53.71	<.0001
	4	1.5960	0.35		
	8	1.7207	0.18		
	10	1.7895			
gender	M	1.6272	0.23	10.10	0.0015
	F	1.6799			
race	Majority	1.6671	0.12	1.22	0.2693
	Minority	1.6399			
diagnosis	ND	1.7267	0.64	61.66	<.0001
	LI	1.5803			
maternal education	<13	1.5793	0.33	13.96	<.0001
	13–16	1.6458			
	17+	1.7354	0.53		
income		0.0144		6.81	0.0091
word				72.25	<.0001
nonverbal IQ	Low	1.6337	0.16	3.91	0.0481
	High	1.6733			
grade x diagnosis	2 ND	1.5602	0.34	1.33	0.2611
	2 LI	1.4554			
	4 ND	1.6741	0.55		
	4 LI	1.5179			
	8 ND	1.8054	0.53		
	8 LI	1.6361			
	10 ND	1.8671	0.49		
	10 LI	1.7120			
grade x nonverbal IQ	2 Low	1.4918	0.10	1.42	0.2343
	2 High	1.5238			
	4 Low	1.5953	0.01		

Effect	Level	Mean Estimate	Effect Size	F Value	Pr > F
	4 High	1.5968			
	8 Low	1.6849	0.22		
	8 High	1.7566			
	10 Low	1.7628	0.16		
	10 High	1.8162			

Note: The Mean Estimate column gives the estimated dependent variable value for categorical variables and the slope for continuous variables (income) computed assuming an average value of every other effect. The Effect Size (mean difference divided by a pooled standard deviation) column compares 2 levels of the effect. The F test is an overall test for the effect.