Journal of Education for Students Placed at Risk (JESPAR)

ISSN: 1082-4669 (Print) 1532-7671 (Online) Journal homepage: https://www.tandfonline.com/loi/hjsp20

Generating Vocabulary Knowledge for At-Risk Middle School Readers: Contrasting Program Effects and Growth Trajectories

Joshua F. Lawrence, Rebecca Givens Rolland, Lee Branum-Martin & Catherine E. Snow

To cite this article: Joshua F. Lawrence, Rebecca Givens Rolland, Lee Branum-Martin & Catherine E. Snow (2014) Generating Vocabulary Knowledge for At-Risk Middle School Readers: Contrasting Program Effects and Growth Trajectories, Journal of Education for Students Placed at Risk (JESPAR), 19:2, 76-97, DOI: 10.1080/10824669.2014.958836

To link to this article: https://doi.org/10.1080/10824669.2014.958836

Published online: 06 Nov 2014.

Journal of Education for Students Placed at Risk, 19: 76–97, 2014

ISSN: 1082-4669 print / 1532-7671 online
DOI: 10.1080/10824669.2014.958836

Routledge
Taylor & Francis Group

Generating Vocabulary Knowledge for At-Risk Middle School Readers: Contrasting Program Effects and Growth Trajectories

Joshua F. Lawrence
University of California, Irvine

Rebecca Givens Rolland
Harvard Graduate School of Education

Lee Branum-Martin
Georgia State University

Catherine E. Snow
Harvard Graduate School of Education

We tested whether urban middle-school students from mostly low-income homes had improved academic vocabulary when they participated in a freely available vocabulary program, Word Generation (WG). To understand how this program may support students at risk for long-term reading difficulty, we examined treatment interactions with baseline achievement on a state standardized test and also differential effects for students with (n = 398) and without (n = 1,395) individualized education plans (IEPs). Students in this unmatched quasi-experiment (5 WG and 4 comparison schools) completed pre- and postvocabulary assessments during the intervention year. We also retested student vocabulary knowledge after summer vacation and the following spring on 11 target words to construct a longitudinally consistent scaled score across 4 waves of data. Growth models show that students experienced summer setback. Although there were no average underlying differences in growth or differences in summer setback for students by baseline achievement, better readers improved more from program participation. IEP status did not predict differential benefits of program participation, and students with IEPs maintained gains associated with participation in WG; however, participation in the program did not change underlying growth trajectories favoring students who did not have IEPs.

Although vocabulary is recognized as a key component of skilled reading (National Institute of Child Health and Human Development [NICHD], 2000), there have been few longitudinal evaluations of vocabulary interventions in urban school districts, which in the United States have higher average levels of poverty and lower reading scores than the national average (National

Correspondence should be addressed to Joshua F. Lawrence, Department of Education, University of California, Irvine, 3200 Education Building, Irvine, CA 92697-5500. E-mail: jflawren@uci.edu

Center for Educational Statistics, 2013). Vocabulary skill develops slowly through recurrent encounters with rich language in discussion and reading, and is strongly influenced by home and neighborhood language exposure. Students who have well-developed reading abilities tend to read more often and have more developed capacities to learn the meanings of words encountered in text or during classroom instruction (Ceci & Papierno, 2005), in addition to having more regular interaction with academic words than their less proficient peers (Stanovich, 1986). Clear individual differences exist in students' abilities both to learn (Fukkink, Blok, & de Glopper, 2001; Fukkink & de Glopper, 1998) and to remember instructed vocabulary (Burns & Boice, 2009). In this article, we use multilevel longitudinal models to describe the vocabulary-learning trajectories of students with higher and lower baseline achievement scores, some of whom have been placed on individualized educational programs (IEPs). We also describe the long- and short-term impact of student participation in a freely available cross-content program focused on all-purpose academic vocabulary intervention: the Word Generation (WG) program. In our analysis we explore whether or not differences exist between students with higher and lower baseline achievement, and between students with ($n = 398$) and without ($n = 1{,}395$) IEPs in their learning and maintenance of instructed vocabulary items during the middle school years.

BACKGROUND AND CONTEXT

Longitudinal Studies of Reading Development Differences Over Time

Longitudinal studies suggest that, without proper intervention, lower-skilled students are likely to fall further behind their more skilled peers in academic domains, a phenomenon that is known as the *Matthew effect* (Morgan, Farkas, & Hibel, 2008; Stanovich, 1986). One study of elementary students' word recognition skills across 3 years found that absolute differences between stronger and weaker students in word recognition increased, although ranking did not shift greatly over time (Bast & Reitsma, 1998). Morgan et al. (2008) found that elementary children who began the study with risk factors for reading failure (such as those from low-socioeconomic [SES] homes or those with low baseline reading skills) gained less than students with higher baseline scores on reading-related skills. On the other hand, Shaywitz et al. (1995) found no evidence for a fanning-out effect of poorer and more skilled readers' scores over time.

Unfortunately, like most longitudinal studies of children's reading and vocabulary development, these studies focus on younger children. There is reason to believe that a Matthew effect might also be evident in middle grades, especially in the development of academic vocabulary. In the following sections, we review research related to individual differences in vocabulary learning, the maintenance of academic skills over the summer, and the research related to targeted vocabulary interventions as a frame for our longitudinal evaluation of an academic vocabulary intervention.

Individual Differences in Vocabulary Learning and Maintenance

Vocabulary is an important component of skilled reading (NICHD, 2000), and increasingly correlates to reading comprehension as students age (Snow, Porche, Tabors, & Harris, 2007).

Educators know a great deal about individual differences in vocabulary learning. Older students and students with higher vocabulary scores are better able to learn new words from texts, suggesting a possible mechanism for a Matthew effect in vocabulary development (McKeown, 1985; Swanborn & de Glopper, 1999; van Daalen-Kapteijns, Elshout-Mohr, & de Glopper, 2001). Skilled readers learn words from context more reliably than their peers matched on vocabulary knowledge (Cain, Oakhill, & Elbro, 2003), although learning differences are reduced in more supportive instructional settings (Cain, Oakhill, & Lemmon, 2004). Given these trends, we are not surprised that adolescent students' abilities to learn words independently, from leisure reading, vary (Lawrence, 2009).

Student maintenance of vocabulary knowledge also varies by instruction and individual learning profile. Older students' knowledge of taught second-language vocabulary tends to regress after instruction (Lovelace & Stewart, 2009; Min, 2008; Zhang & Schumm, 2000), although multiple exposures (Folse, 2006; Rott, 2007) and output-oriented instruction (Kitajima, 2001) may support retention. Ricketts, Bishop, and Nation (2008) found that, although 9- to 10-year-old good and poor comprehenders showed similar ease in learning the meanings of invented words, poor comprehenders did not retain these words as well. In general, although participatory and output-oriented activities allow students to preserve gains longer than passive techniques, more needs to be known about the differential word-learning trajectories of students with and without IEPs, as well as students with stronger and weaker baseline reading skills.

Academic Skills Decline During Summer

There has been research on how well students learn and preserve academic skills during the summer months (Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996; Entwisle, Alexander, & Olson, 2000; Heyns, 1978; Kim, 2004), and on differences in students' summer setback by home income status. The impact of summer setback has been interpreted as cumulative: Differences between students with strong and weak literacy skills increase during summer, so even if learning is equivalent during the school year, large differences accrue (Entwisle et al., 2000). Home SES status has been shown to relate to differences in summer setback, but this does not provide much information to educators working in increasingly economically homogeneous urban school districts.

A decision of the U.S. District Court for the Eastern District of Pennsylvania in *Armstrong v. Kline*, 476 F. Supp. 583 (E.D. Pa. 1979), noted the inadequacy of 180 days of schooling for students with severe disabilities, although it did not address the impact of summer setback for students who have mild reading disabilities or are low-achieving readers. There are some studies on summer setback for students with low baseline achievement. Mraz and Rasinski (2007) examined fluency and decoding skills among first-, second-, and third-grade students during summer break and found that summer setback was larger for low-achieving students. Ross (1974) also found summer setback for low-achieving sixth-grade students ($n = 119$), but not for their higher achieving peers. In their meta-analysis of summer learning studies, Cooper et al. (1996) found larger summer setback for older students than for younger students. Together these studies suggest that lower-achieving middle-grade students may be particularly susceptible to summer setback; more needs to be done to understand these differences within low-SES urban districts.

Vocabulary Interventions Improve Reading Comprehension

Vocabulary interventions can help students improve their reading skills. When students receive rich vocabulary instruction that includes multiple contextualized exposures to target words, they not only learn new words, but can also understand passages that contain those target words better (McKeown, Beck, Omanson, & Perfetti, 1983). In a meta-analysis of vocabulary instruction methods, Stahl and Fairbanks (1986) concluded that rich vocabulary instruction improved passage comprehension, when taught words both were and were not included in the tested passages. Vaughn et al. (2009) found that students who took part in a 12-week social studies vocabulary intervention, including explicit instruction, peer discussion, and graphic organizers, improved on measures of vocabulary and comprehension. Other researchers have similarly integrated such interventions into content-area classes, finding benefits for strategy instruction (Winchester et al., 2009) and target words taught using principles of rich vocabulary instruction (Townsend & Collins, 2009). A randomized trial of the Quality English and Science Teaching program found that participating students improved in both vocabulary knowledge and science-content knowledge (August, Branum-Martin, Cardenas-Hagan, & Francis, 2009), and Bos and Anders (1990) found that content-area vocabulary instruction resulted in improved content-area reading ability for students identified as having learning disabilities. In summary, there is evidence that vocabulary instruction can be infused into content-area classes, improving both vocabulary knowledge and disciplinary knowledge.

The WG program is a free schoolwide vocabulary intervention designed around the principles of effective vocabulary instruction (Strategic Educational Research Partnership, 2009). The program focuses on five target words each week and is delivered in each content-area class (i.e., English language arts, mathematics, science, and social studies) once a week, so that each teacher teaches the curriculum for 15 min per week to each class, and the students receive instruction in the words each day in a different content-area class. Each Monday, teachers introduce students to five target words that are embedded in a short passage about a social dilemma. These passages are designed to engage students, requiring them to take a position about an issue. For instance, some units deal with whether a practice should be legal (e.g., renting a pet or wiretapping); others may deal with an issue more immediately relevant to middle-school students (e.g., whether or not single-gender education is a good idea). On each of the following weekdays (Tuesday to Thursday) students engage in a short WG activity, each day in a different content area (mathematics, science, or social studies). In these classes they reencounter the target words in a 15-min instructional session based around a discipline-appropriate examination of the weekly topic. On Friday, they write an essay stating their position on the topic in an English language arts class, as informed by their readings and class discussion throughout the week.

Previous research on WG (Snow, Lawrence, & White, 2009) suggests that students in the program exhibited greater growth on target vocabulary than those who did not participate, and that improvement on target-word knowledge predicted improved performance on the Massachusetts Comprehensive Assessment System (MCAS) English language arts assessment. A second study examined how well students from language-minority homes learned and maintained the word knowledge associated with program participation (Lawrence, Capotosto, Branum-Martin, White, & Snow, 2012). These studies were conducted in large urban districts with large numbers of students from low-SES homes, but we have not examined differential benefits for students with varying levels of baseline achievement scores or by IEP status.

Special Education Students Benefit from Vocabulary Instruction

Several studies have focused on the effectiveness of vocabulary interventions targeted specifically for adolescents with reading difficulties (e.g., Seifert & Espin, 2012; Wanzek, Vaughn, Roberts, & Fletcher, 2011), with one study using a within-subject design showing a text-reading intervention had benefits for students' reading fluency and vocabulary (Seifert & Espin, 2012). One form of intervention is teaching the analysis of graphosyllabic units as a means of improving vocabulary knowledge. Studies have found that this method benefits students with low reading skills (Bhattacharya & Ehri, 2004) and helps students learn new vocabulary incidentally from textual encounters (Penno, Wilkinson, & Moore, 2002), although the latter study notably found that students with high initial reading skills made greater gains in vocabulary than those with low initial skills. Roberts, Torgesen, Boardman, and Scammacca (2008) found that direct instruction helped students with learning disabilities acquire vocabulary from reading texts at their independent and instructional reading levels. Similarly, Burns and Boice (2009) replicated this finding by showing that students with learning disabilities who had the greatest opportunity to practice target words had the greatest gains and retention of those words 1 to 2 weeks later. Bryant, Goodwin, Bryant, and Higgins's review (2003), which focused on studies of instruction for students with learning disabilities, found that instruction that used mnemonic imagery was more effective than traditional direct instruction (Mastropieri, Scruggs, Levin, Gaffney, & McLoone, 1985). Bos and Anders (1990) reported that instruction including semantic mapping and semantic feature analysis produced better long-term recall than definitional instruction alone. Using a multiple baseline by groups design, Bulgren, Schumaker, and Deshler (1988) found an increase in vocabulary from concept-focused instruction for students in both special and general education; Fore, Boon, and Lowrie's (2007) six learning-disabled participants improved more in a concept-model condition than in the definition mode of instruction. Such research, as well as research on root-word analysis and metalinguistic awareness (Ebbers & Denton, 2008), identifies promising instructional techniques and highlights the need to develop interventions that allow students with low initial skills to catch up with their higher-skilled peers.

Although these findings suggest that students on IEPs can benefit from targeted vocabulary intervention, interventions are often resource intensive. There is an increasing tendency for secondary-school classrooms to include students with IEPs (Downing & Peckham-Hardin, 2007; McLeskey & Henry, 1999; McLeskey, Henry, & Hodges, 1999), which in some cases may allow for more flexible use of available teachers in schools with large numbers of students at risk for reading difficulty (with or without IEPs). Research suggests some support for vocabulary in embedded classrooms: Older children can learn and generalize vocabulary in these settings (McDonnell et al., 2006), and inclusion can promote peer tutoring between learning-disabled and general education students (Hughes & Fredrick, 2006). To date, the long-term learning trajectories of students in response to schoolwide vocabulary instruction have not been described, nor have differences in learning trajectories been related to baseline achievement.

Although vocabulary and reading comprehension are strongly related (Snow et al., 2007) and many students with IEPs have significant weaknesses in vocabulary (Hock et al., 2009), no longitudinal study of academic vocabulary learning for students with and without IEPs and with high and low baseline scores has been conducted. There is no research on how a modest and free curricular intervention like WG may influence students' short- and long-term academic vocabulary knowledge. Therefore, our research questions are:

1. How did high achieving students without IEPs who participated in the WG program learn, maintain, and consolidate words compared to similar students attending comparison schools?
2. How did students with IEPs who participated in the WG program learn, maintain, and consolidate words compared with similar students attending comparison schools?
3. How did students with low baseline achievement scores on the state's standardized English language arts assessment who participated in the WG program learn, maintain, and consolidate words compared to similar students attending comparison schools?

METHODS

We conducted an unmatched quasi-experiment to determine the efficacy of the WG program. Treatment schools implemented the WG program, providing daily instruction on five target words each week (as described in the following). We collected pre- and posttest data during the first year of the quasi-experiment from five treatment schools and four comparison schools. Students in comparison schools received varying degrees of vocabulary instruction, depending on the school. We know, for instance, that vocabulary was a focus of intense professional development in one school, but we do not have systematic observations of work done in either treatment or control schools. None of the comparison schools had cross-content programs like WG to support their students' vocabulary development. During the following year, the WG schools implemented a second year of the program, with a new set of topics and target words. We embedded a subset of 11 year-1 items in the year-2 assessment to determine how well students retained knowledge of words taught during the first year.

District Setting

The WG program was developed in partnership with Boston Public Schools to meet needs specified by the district early in the partnership. Prior to the start of this quasi-experiment, early versions of WG materials were piloted in two schools and modified based on feedback from the teachers; this work was essential in refining and improving the program. The district agreed to participate in a quasi-experiment of WG. WG schools volunteered to implement the program to address a perceived weakness in their curriculum; comparison schools were selected by the district. The fact that treatment schools identified themselves as wanting vocabulary support might explain why the comparison schools performed better on baseline state assessments and on our vocabulary assessment than treatment schools did.

School Sites

The three right-hand columns in Table 1 report the percentage of students in the warning category on the English language arts MCAS by grade level in each school, and shows that a

TABLE 1
Number of Students Eligible for Federally Funded Free and Reduced Lunch Program, with Individualized Education Plans, and in the Warning Proficiency Category on Massachusetts Comprehensive Assessment System (MCAS) for All Treatment and Control Schools

	Eligible for Federally Funded Free and Reduced Lunch		*Students with IEP*		*Percentage of Students in the Warning Category in English Language Arts MCAS*		
School	*Percentage*	*Total*	*Percentage*	*Total*	*6th*	*7th*	*8th*
Treatment							
Reilly	83.6	320	26.9	103	4.9	6.7	2.5
Mercer	87.7	415	19.5	92	9.9	5.8	11.7
Westfield	88.6	150	32.5	55	20.3	19.6	11.4
Mystic	90.6	125	13.0	18	2.5	7.1	0
Occidental	87.4	90	31.1	32	19.5	29	25
Comparison							
Walters	76.7	69	14.4	13	2.6	0	0
Garfield	80.4	45	28.6	16	0	0	5.9
Jefferson	79.2	84	30.2	32	11	5.6	n.a.
Uxton	86.0	243	13.1	37	5.4	8.3	2
Average							
Treatment	86.4	1090	23.8	300	10.2	9.8	8.7
Comparison	83.2	433	18.8	98	6.4	4.9	2.2

Note. IEP = Individualized education plan.

higher proportion of readers were in the warning category in the treatment schools at the beginning of the study. Furthermore, treatment schools had more students with IEPs (23.8%) than the comparison schools (18.8%), although roughly similar numbers of students were eligible for free and reduced-price lunch in the treatment (86.4%) and control (83.2%) conditions.

Student Sample

A districtwide code was provided for each student to identify the range of services for which he or she qualified under state and federal regulations (Table 2). By far the largest group of students with IEPs in the district, and this study, was the group classified as needing resource services. These students attended general education classes and were supported by an additional special education teacher or paraprofessional in math or English classes. Smaller numbers of students attended some or all of their classes in substantially separate classrooms. At several schools, the special education teachers taught all the WG materials. In other schools, students did not participate in every WG lesson because they did not have the same academic schedule as their peers. Because of the small numbers of students in each subgroup (Table 2), we could not complete a separate analysis for each student designation, but instead completed the analysis on the basis of whether students had an IEP or not (irrespective of the specific level of support services they were provided). Because all

TABLE 2
Service Categories Described by Individualized Education Plans of All Students

Services	*N*	*Percentage*
No IEP	1,395	77.8
Resource services	214	11.9
Language based classroom	29	1.6
Social academic remediation	51	2.8
Learning disabled	73	4.0
Moderate to severe emotional or behavioral disability	19	1.1
Other	13	0.7

Note. IEP = Individualized education plan.

of these schools were middle schools, during the second year (the follow-up year) the students in the oldest cohort graduated to high school and thus are not included in the follow-up analysis.

Procedure

During the first academic year of the quasi-experiment (Fall [w1] to Spring [w2]), students in the treatment schools participated in the WG program. To examine summer setback in students' vocabulary knowledge, we embedded 11 items from the first year of instruction into the pre- and posttests. We constructed a longitudinally consistent measure to maximize information from these 11 items using an item response theory (IRT) approach. First, a single factor model was fit to the 11 items at each wave to determine whether the 11 items were good indicators of a single factor of vocabulary knowledge. Second, we used the item parameters from wave 1 to produce scaled scores for each of the three subsequent waves. We provide details about the scaling in the measures section.

The district provided the IEP status and MCAS information for most of the students who completed either the pre- or posttest during the instructional year. Unfortunately, data about the IEP status ($n = 174$) or MCAS ($n = 22$) were missing for some students; we dropped these students from the analytical sample so that competing models could be compared using the model deviance. We found no difference between dropped and retained students on baseline MCAS or vocabulary scores. Even though we had to remove students with missing predictor data from the analysis, we did use data contributed by students who completed as little as a single pre- or posttest during the instructional year because longitudinal methods do not require that subjects contribute more than one wave of outcome data (Singer & Willett, 2003). Table 3 provides a description of the resulting analytical sample. Scanning down the rows suggest that less data was collected in each subsequent wave of data collection. Student movement across districts and between schools is responsible for most of the attrition between waves 1 and 2 and waves 3 and 4. The high attrition between waves 2 and 3 reflects the fact that the oldest cohort of students (those who started the study as eighth graders) had graduated to high school by the third wave of data collection and were unavailable to contribute data to the study after that time.

TABLE 3
Number of Treatment and Comparison School Students Contributing Data at Each Wave of Data Collection

	Number of Students Who Contributed Data at Each Wave		
	Total	*IEP*	*No IEP*
Comparison school students			
Fall (wave 1)	481	88	393
Spring (wave 2)	365	66	299
Fall (wave 3)	174	35	139
Spring (wave 4)	231	48	183
Treatment school students			
Fall (wave 1)	1037	229	812
Spring (wave 2)	1072	258	817
Fall (wave 3)	681	153	529
Spring (wave 4)	617	127	490

Note. IEP = Individualized education plan.

Measures

Outcome: Vocabulary

The target words tested in this study were general academic words that are widely dispersed in texts across content areas, words such as *relevant*, *presume*, and *indicate* (Coxhead, 2000). We collected student performance data on 50 items at each wave of data collection. Each assessment item presented a target word, underlined, in a neutral sentence context; the students chose from among four options the closest synonym for the target word. Student scores on all items were used to establish the IRT model, which formed a time-varying level-1 outcome VOCAB. The IRT scaled score was produced by fitting a single-factor confirmatory factor analysis model to the 11 repeated items separately for each complete wave of vocabulary data, using Mplus 5, with robust weighted least squares estimation for dichotomous data (Muthén & Muthén, 2007; see Table 4). The model fit reasonably well in all four waves, as shown in Table 5. Although there was some degree of misfit in the first wave (CFI = .94), the root mean square error of approximation was quite acceptable for all waves (RMSEA $\leq$ .03). We used the item parameters (loadings and thresholds) from the first wave to score the following three waves, thereby estimating a factor score on the metric of the first wave, with factor means and variances free to differ over time. In this way, the vocabulary scores

TABLE 4
Fit Statistics for Categorical CFA Models for Each Wave

Wave	*Chi-square (df)*	*CFI*	*RMSEA*	*WRMR*
1	111.1 (41)	0.941	0.031	1.21
2	100.4 (42)	0.962	0.028	1.12
3	125.1 (42)	0.958	0.027	1.25
4	67.1 (42)	0.99	0.016	0.9

Note. CFI = Comparative fit index. RMSEA = Root mean square error of approximation. WRMR = Weighted root mean square residual. All models fit with robust weighted least squares estimation (WLSMV; Muthén & Muthén, 2007).

TABLE 5
Scaled and Raw Vocabulary Scores for All Treatment and Comparison Schools at Each Wave of Data Collection

	Scaled				*Raw*			
	Instructional Year		*Follow Up Year*		*Instructional Year*		*Follow Up Year*	
School	*Fall W1*	*Spring W2*	*Fall W3*	*Spring W4*	*Fall W1*	*Spring W2*	*Fall W3*	*Spring W4*
Comparison								
Garfield								
Mean	0.20	0.47	n.a.	0.37	5.67	5.94	n.a.	6.36
SD	(0.71)	(0.85)	n.a.	(0.78)	(2.17)	(2.35)	n.a.	(2.10)
N	51	52	0	36	51	52	0	36
Jefferson								
Mean	0.15	0.40	−0.01	0.24	5.38	5.53	5.34	5.89
SD	(0.84)	(0.96)	(0.81)	(0.98)	(2.40)	(2.69)	(2.49)	(2.83)
N	98	103	57	54	98	103	57	54
Uxton								
Mean	0.29	0.70	0.32	0.78	5.86	6.58	6.22	7.39
SD	(0.76)	(0.83)	(0.77)	(0.78)	(2.19)	(2.21)	(2.27)	(2.19)
N	242	210	117	141	242	210	117	141
Walters								
Mean	0.23	n.a.	n.a.	n.a.	5.69	n.a.	n.a.	n.a.
SD	−0.69	n.a.	n.a.	n.a.	−2.05	n.a.	n.a.	n.a.
N	90	0	0	0	90	0	0	0
Average								
Mean	0.24	0.58	0.21	0.59	5.71	6.19	5.93	6.88
SD	(0.76)	(0.88)	(0.80)	(0.86)	(2.21)	(2.42)	(2.38)	(2.42)
N	481	365	174	231	481	365	174	231
Treatment								
Reilly								
Mean	−0.06	0.54	0.15	0.47	4.72	6.29	5.83	6.77
SD	(0.73)	(0.77)	(0.77)	(0.82)	(2.14)	(2.10)	(2.28)	(2.46)
N	312	349	208	192	312	349	208	192
Mercer								
Mean	−0.02	0.48	0.15	0.52	4.92	5.95	5.82	6.66
SD	(0.76)	(0.85)	(0.81)	(0.79)	(2.21)	(2.43)	(2.39)	(2.26)
N	429	359	253	244	429	359	253	244
Westfield								
Mean	−0.15	0.22	−0.13	0.34	4.59	5.20	4.88	6.21
SD	(0.70)	(0.79)	(0.83)	(0.85)	(2.01)	(2.23)	(2.42)	(2.62)
N	94	142	94	62	94	142	94	62
Mystic								
Mean	0.02	0.62	0.22	0.59	4.97	6.27	5.91	6.98
SD	(0.71)	(0.78)	(0.72)	(0.85)	(2.05)	(2.13)	(2.06)	(2.40)
N	127	135	86	85	127	135	86	85
Occidental								
Mean	−0.29	0.19	−0.31	0.11	4.08	5.37	4.19	5.66
SD	(0.68)	(0.87)	(0.65)	(0.83)	(2.00)	(2.49)	(2.19)	(2.46)
N	77	88	41	35	77	88	41	35
Average								
Mean	−0.06	0.46	0.09	0.47	4.77	5.95	5.61	6.64
SD	(0.76)	(0.88)	(0.80)	(0.86)	(2.21)	(2.42)	(2.38)	(2.42)
N	1037	1072	681	617	1037	1072	681	617

for each wave were estimated on a single, consistent metric, relative to the first wave. Thus, the 11 items generate a reliable and consistent score across waves.

Question Predictors

Wave. WAVE is an individual (level-1) variable indicating wave of data collection (0 through 3).

Instruction. INSTRUCTION is a time-varying individual (level-1) variable that indicates if students have had an instructional encounter with the target words. Because students in WG schools were instructed on these target words during the first, but not the second, year, the variable for those students is coded as follows: wave 0 = *0*, wave 1 = *1*, wave 2 = *1*, wave 3 = *1*. Comparison-school students were not systematically instructed on these words, so INSTRUCTION was coded as *0* for them at all waves.

Summer. SUMMER indicates the number of summers students had experienced since the start of the study (wave 0 = *0*, wave 1 = *0*, wave 2 = *1*, wave 3 = *1*). It is a time-varying individual (level-1) variable.

Attends a WG school. WG_SCHOOL indicates if students attended a WG school (WG_SCHOOL = *1*) or a comparison school (WG_SCHOOL = *0*). It is a level-2 variable.

MCAS. MCAS is a level-2 variable recording each student's baseline score on the state standardized assessment of English language arts, which has been *z*-score transformed by subtracting the mean score from each score and dividing by the standard deviation.

IEP. IEP is a level-2 variable indicating if a student was on an individualized education program (IEP = *1*) or not (IEP = *0*).

Grade-level cohort. All students began the study in grade 6, 7, or 8. Grade level was provided by the school district and used to create two variables. GRADE7 describes if the student was in seventh grade (GRADE7 = *1*) or not (GRADE7 = *0*). GRADE8 describes if the student was in eighth grade (GRADE8 = *1*) or not (GRADE8 = *0*).

Analysis

We used the multilevel model for change (Singer & Willett, 2003) to address each research question. Due to the limited number of waves of data available, we assumed that growth was linear within the year, and included a parameter for a one-time summer setback. Level-2 variance in the rate-of-change parameter was negligible in all fitted models so it was fixed to zero. All models that were explored in determining the final fitted model were based on a level-1, level-2 model with the following specifications:

Level 1:

$$\widehat{VOCAB} = \pi_{0i} + \pi_{1i} WAVE_{ij} + \pi_{2i} INSTRUCTION_{ij} + \pi_{3i} SUMMER_{ij} + \varepsilon_{ij} \quad (1)$$

Level 2:

$$\pi_{0i} = \gamma_{00} + \gamma_{01} GRADE7_i + \gamma_{02} GRADE8_i + \gamma_{03} WG\_SCHOOL_i + \gamma_{04} IEP + \gamma_{05} MCAS$$

$$\begin{aligned}
\pi_{1i} &= \gamma_{10} + \gamma_{11} GRADE7_i + \gamma_{12} GRADE8_i + \gamma_{13} WG_SCHOOL_i + \gamma_{14} IEP_i \\
&\quad + \gamma_{15} IEP_i \times \mathrm{W}G_SCHOOL_i + \gamma_{16} MCAS \\
\pi_{2i} &= \gamma_{20} + \gamma_{21} GRADE7_i + \gamma_{22} GRADE8_i + \gamma_{23} IEP_i + \gamma_{24} MCAS \\
\pi_{3i} &= \gamma_{30} + \gamma_{31} WG_SCHOOL_i + \gamma_{32} IEP_i + \gamma_{33} IEP_i \times \mathrm{W}G_SCHOOL_i + \gamma_{34} MCAS
\end{aligned}$$

where $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$.

Specific estimates of parameters in the final fitted model were used to answer each research question. To answer the first research question, which asks about vocabulary learning trajectories of students in treatment and control schools, we examine the estimated impact of instruction (γ_{20}) and any interactions between instruction and grade level (γ_{21}, γ_{22}). To answer question two, which focuses on differences in trajectories according to student IEP status, we examine the intercept, slope, summer setback and the impact of instruction for both students without IEPs ($\gamma_{00}, \gamma_{10}, \gamma_{30}, \gamma_{20}$) and those with IEPs ($\gamma_{04} IEP_i, \gamma_{14} IEP_i, \gamma_{33} IEP_i, \gamma_{23} IEP_i$). To answer question three, which focuses on differences in trajectories according to student MCAS scores, we will examine how intercept, slope, summer setback and the impact of instruction vary according to MCAS performance ($\gamma_{05} MCAS, \gamma_{16} MCAS, \gamma_{24} MCAS, \gamma_{34} MCAS$).

RESULTS

Table 5 provides the average raw and scaled vocabulary scores for all schools at each wave of data collection. Comparing the first and second data column on the left side shows that students in every school improved during the experimental year (Fall [w1] to Spring [w2]). Comparing the second data column with the third (Spring [w2] to Fall [w3]) suggests that students in all schools regressed in their academic word knowledge during the summer, but then improved during the final year, when school was in session again (Fall [w3] to Spring [w4]). This table suggests that comparison-school students improved in knowledge of target words during both academic years, suggesting that students learn these words when attending middle schools that do not participate in an explicit cross-content vocabulary program. This table also suggests that students in the treatment schools improved more ($M_{wave4} - M_{wave1} = 0.53$, scaled score) than students in the comparison schools ($M_{wave4} - M_{wave1} = 0.35$, scaled score) on average, and that their relative improvement occurred in the first year of the study.

The top half of Table 6 presents the average scaled and raw vocabulary scores of students with and without IEPs in the treatment and comparison schools. Comparing the first and fourth data column for each subgroup allows us to calculate students' improvement in the scaled metric from the beginning of the study to the end of the study. Students in the comparison group

TABLE 6
Scaled and Raw Vocabulary Scores for Students With and Without IEPs and With Above or Below Average Baseline MCAS Scores at Each Wave of Data Collection

	Scaled Score				*Raw Score*			
	Instructional Year		*Follow Up Year*		*Instructional Year*		*Follow Up Year*	
	Fall Wave 1	*Spring Wave 2*	*Fall Wave 3*	*Spring Wave 4*	*Fall Wave 1*	*Spring Wave 2*	*Fall Wave 3*	*Spring Wave 4*
Comparison								
IEP	−0.15	−0.02	−0.18	0.03	4.73	4.38	4.64	5.27
	(0.72)	(0.82)	(0.69)	(0.86)	(2.20)	(2.42)	(2.07)	(2.39)
No IEP	0.33	0.72	0.31	0.74	5.93	6.61	6.27	7.31
	(0.74)	(0.84)	(0.79)	(0.80)	(2.15)	(2.22)	(2.34)	(2.25)
Treatment								
IEP	−0.51	−0.07	−0.46	−0.13	3.58	4.61	4.01	4.95
	(0.66)	(0.75)	(0.73)	(0.77)	(1.94)	(2.17)	(2.14)	(2.29)
No IEP	0.07	0.62	0.25	0.62	5.11	6.37	6.06	7.07
	(0.71)	(0.77)	(0.73)	(0.76)	(2.08)	(2.17)	(2.21)	(2.24)
Comparison								
Low MCAS	−0.23	0.12	−0.16	0.17	4.40	4.94	4.86	5.65
	(0.61)	(0.82)	(0.78)	(0.79)	(1.86)	(2.30)	(2.31)	(2.22)
High MCAS	0.51	0.87	0.59	0.90	6.46	7.02	7.02	7.79
	(0.70)	(0.78)	(0.60)	(0.76)	(2.03)	(2.10)	(1.89)	(2.14)
Treatment								
Low MCAS	−0.40	0.03	−0.27	0.07	3.87	4.75	4.51	5.44
	(0.64)	(0.72)	(0.70)	(0.78)	(1.87)	(2.02)	(2.00)	(2.24)
High MCAS	0.19	0.80	0.43	0.81	5.45	6.92	6.62	7.66
	(0.69)	(0.72)	(0.71)	(0.69)	(2.07)	(2.03)	(2.21)	(2.05)

Note. MCAS = Massachusetts Comprehensive Assessment System. IEP = Individualized education plan.

who had IEPs started the program with relatively low scores, and made only modest gains ($M_{wave4} - M_{wave1} = 0.18$). Students with IEPs in the treatment group began the study with even lower baseline vocabulary scores ($M_{\mathrm{wave1}} = -0.51$), but improved 0.38 scaled points by the spring of this follow-up year. Students without IEPs improved more than those with them in both comparison ($M_{wave4} - M_{wave1} = 0.41$) and treatment schools ($M_{wave4} - M_{wave1} = 0.55$). These trends suggest that students without IEPs are pulling ahead of students with them. Interestingly, the difference between students with and without IEPs was similar in both the treatment and comparison schools across the four waves of data collection. This suggests that differences in academic vocabulary growth across the study may be associated with differences in underlying growth trajectories rather than differences in response to the WG program.

The bottom half of Table 6 displays the average scaled and raw vocabulary scores for students with above average or below average scores on the MCAS assessment in comparison and treatment schools. Again, we find differences in improvement across subgroups. Students with below average MCAS scores improved ($M_{wave4} - M_{wave1} = 0.40$) roughly as much as students with above average MCAS scores ($M_{wave4} - M_{wave1} = .39$) in the comparison schools. In contrast, there was a pronounced difference in the improvement between students with below average ($M_{wave4} - M_{wave1} = .47$) and above average ($M_{wave4} - M_{wave1} = .62$) MCAS scores in the treatment schools. Students with IEPs had much lower MCAS scores than their peers without them (one standard deviation lower on average). Although these descriptive tables help us to anticipate some of the trends we might expect to see in our longitudinal growth models, they do not control for multiple student-level factors nor do they take advantage of the multiple waves of data contributed by each student. Therefore, we use multilevel models to answer our research questions.

1. How Did High-Achieving Students Without IEPs Who Participated in the WG Program Learn, Maintain, and Consolidate Words, Compared to Similar Students Attending Comparison Schools?

To understand the vocabulary growth trajectories of students in the WG schools and the comparison schools, we fit a series of multilevel models predicting students' scaled vocabulary scores (Table 7). Model A is the unconditional growth model except that it also includes an important time-varying predictor: the number of summers experienced. This model demonstrates that students tended to improve in their vocabulary knowledge at each successive assessment, except during the summer when they experienced a summer setback (γ_{30} = –0.712, p < .001). Model B is the final fitted model. The estimates of key coefficients in this model show that students improved at each wave of assessment (γ_{10} = 0.373, p < .001) except during summer when the coefficient for having experienced a summer (γ_{30} = 0.610, p < .001) was greater than the underlying wave-to-wave improvement. Students in the treatment schools start off the study with lower average scaled vocabulary scores than students in the comparison schools (γ_{03} = –0.256, p < .001). Students with IEPs have lower average scaled vocabulary scores than students who do not have them (γ_{04} = –0.154, p < .001). Similarly, baseline MCAS scores predict higher vocabulary scores such that a one standard-deviation difference in MCAS scores is equal to a 0.354 difference in baseline scaled vocabulary on average (γ_{05} = 0.354, p < .001). Average baseline vocabulary scores of students in seventh (γ_{01} = 0.122, p < .001) and eighth (γ_{02} = 0.366, p < .001) grade are higher than those of sixth-grade students.

We also found that students in the treatment group improved more during the instructional period than students in the comparison school (γ_{20} = .207, p < 0.001). We explored, but did not find, interactions between treatment and growth across the four waves of data collection; improvement associated with program participation only occurred during the instructional year. We also explored interactions between grade level and WG participation and found that students in eighth grade did not improve as much as students in grades 6 and 7 when participating in the intervention (γ_{22} = –0.125, p < 0.001). These results are demonstrated by the bold and the light dashed lines in Figure 1 (the top two lines). These lines chart the trajectories of prototypical sixth-grade students with average MCAS scores who do not have IEPs. Students in the treatment schools (represented by the bold dashed line) made significant gains compared to students in the comparison schools (represented by the light dashed line) between the first and second measurement occasion when they were participating in the program. During the summer, students in both treatment and control schools experienced a setback in their average academic vocabulary scores, which improved again during the following school year. During this time, students who participated in the WG program maintained their improved vocabulary scores relative to the comparison school students; treatment effects were not attenuated.

TABLE 7
Multilevel Models for Change Predicting VOCAB Across Four Waves of Data

	Research Question	*Model A: Unconditional Growth Model Plus Summer*	*Model B: Final Fitted Model*
Level 1 Predictors			
Intercept (γ_{00})		0.0389^{*}	0.0819^{*}
		(0.0191)	(0.0320)
WAVE (γ_{10})		0.436^{***}	0.373^{***}
		(0.0149)	(0.0191)
SUMMER (γ_{30})		-0.712^{***}	-0.610^{***}
		(0.0339)	(0.0364)
Level 2 Predictors			
WG_SCHOOL (γ_{03})			-0.256^{***}
			(0.0334)
IEP (γ_{04})			-0.154^{***}
			(0.0407)
MCAS (γ_{05})			0.354^{***}
			(0.0181)
GRADE7 (γ_{01})			0.122^{***}
			(0.0303)
GRADE8 (γ_{02})			0.366^{***}
			(0.0370)
Question Predictors			
INSTRUCTION (γ_{20})	RQ 1		0.207^{***}
			(0.0285)
GRADE8 X INSTRUCTION (γ_{22})	RQ 1		-0.125^{**}
			(0.0444)
IEP × WAVE (γ_{23})	RQ 2		-0.0416^{*}
			(0.0182)
MCAS × INSTRUCTION (γ_{24})	RQ 3		0.0850^{***}
			(0.0171)
Level 1 Variance Component			
Residual		0.415^{***}	0.200^{***}
		(0.00887)	(0.00512)
Level 2 Variance Component			
Initial Status		0.236^{***}	0.228^{***}
		(0.00312)	(0.00298)
N		4659	4659
Deviance (-2 LL)		9446.6	8376.1

Note. Standard errors in parentheses. MCAS = Massachusetts Comprehensive Assessment System. IEP = Individualized education plan.

$^{*}p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

2. How Did Students with IEPs Who Participated in the WG Program Learn, Maintain, and Consolidate Words Compared with Similar Students Attending Comparison Schools?

To answer research question two, we looked for interactions between students' special-education status and instruction, growth, and summer setback. We found that although there were no

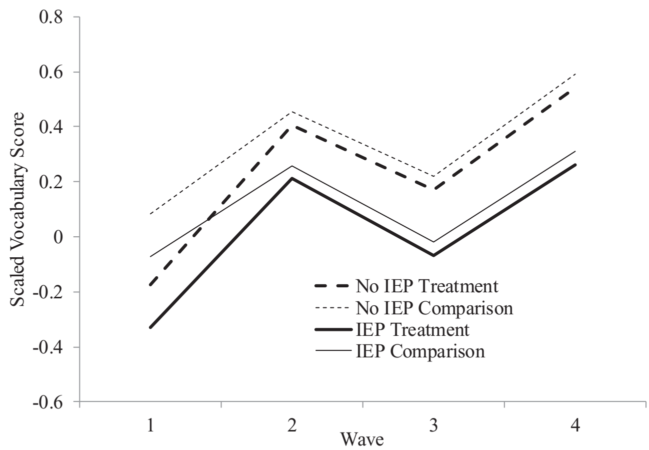

FIGURE 1 Fitted trajectories of sixth-grade students with and without individualized education programs attending treatment or comparison schools.

interactions between special-education status and treatment, there were differences in the underlying growth trajectories of students who did and did not have IEPs ($\gamma_{23} = -0.0416$, $p < 0.05$). This finding can be seen in Figure 1. The solid lines represent the trajectories of prototypical sixth-grade students with average MCAS scores who do have IEPs in the comparison (light line) and treatment schools (bold line). Differences between students with and without IEPs can be seen by comparing the prototypical trajectories of students with IEPs in the comparison schools (the light solid line) with the prototypical trajectory of students without IEPs in the comparison schools (the light dashed line). The difference between these trajectories increases from wave to wave, even though neither experiences the instructional treatment. A similar underlying difference explains the fitted trajectories of prototypical treatment students with IEPs (bold solid lines) and without them (bold dashed lines). Students in both these groups made marked progress relative to their comparison cohort, but differences between them actually grew across the four waves of data collection. Participation in WG was just as effective for students with IEPs as for those without them, although it did not change underlying differences between the growth trajectories of the two groups.

3. How Did Students with Low Baseline Achievement Scores on the State's Standardized English Language Arts Assessment Who Participated in the WG Program Learn, Maintain, and Consolidate Words Compared to Similar Students Attending Comparison Schools?

To answer research question three, we explored interactions between students' MCAS scores and their underlying vocabulary growth, summer setback, and improvement associated with participation in the WG program. We found that in addition to the main effect of program participation described previously, there was also an interaction with standardized MCAS scores so that students with higher MCAS scores improved more from instruction than did students with lower scores ($\gamma_{24} = 0.0850$, $p < 0.001$). These trends can be seen in the prototypical trajectories plotted on Figure 2 (high scores = +0.5 SD, low scores = –0.5 SD). Differences between the trajectories of prototypical sixth-grade students without IEPs with low MCAS scores in the treatment (dashed light line) and comparison schools (dashed bold line) attenuate during the instructional period. However, the differences between scaled vocabulary scores of treatment students (bold solid line) and comparison students (light solid line) with high MCAS scores is completely eliminated on average, due to the strong treatment effect for students with high MCAS scores. There are no differences in the underlying trajectories related to MCAS status, but students with high MCAS scores benefited more from participation in the WG program than those with lower baseline scores.

DISCUSSION

Vocabulary increasingly relates to children's reading comprehension outcomes as they get older (Snow et al., 2007) and to student performance on standardized state assessments (Snow et al., 2009). One of the central objectives of this study was to understand the effects of a vocabulary intervention on students' vocabulary-learning trajectories. We found that, on average, students did benefit from program participation. Students with IEPs benefited just as much as students without IEPs from participation in the WG program. Our findings suggest that providing multiple exposures to target words in a rich, stimulating context is an appropriate way to help students with and without IEPs learn high-leverage academic words. Additionally, students' improvements made during participation in the program lasted 1 year after the instruction. Program participation did not result in a change in student vocabulary learning rates; that is, participating students did not continue to show improvement in their vocabulary knowledge relative to comparison students after the instructional period had ended, but neither did the effects of the program participation attenuate during the following summer or school year relative to

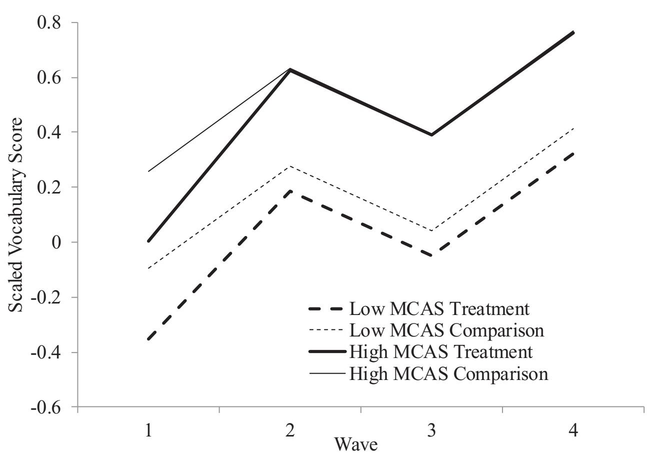

FIGURE 2 Fitted trajectories of sixth-grade students with and without individualized education programs who scored poorly (–.50 SD) or well (+.50 SD) on the baseline English language arts administration of the Massachusetts Comprehensive Assessment System.

comparison students. Similarly, program participation did not change differences in growth across student IEP designation across the 2 years of the study, so although both groups benefited from program participation equally, students without IEPs had greater gains from wave 1 to wave 4.

Although participation in the WG program did not result in changes in students' overall vocabulary learning trajectories, students did achieve one-shot word-learning gains from participation. The gains achieved by participants in the program were significantly larger than the differences associated with IEP status across the 2 years of the study on average. Students with IEPs who participated made significant gains when compared with students who did not participate, and although the relative gains over students without IEPs diminished over time, participating IEP students still had a relative advantage at the end of the study. We take these findings to demonstrate the value of more direct vocabulary instruction that promotes the use of target words in multiple contexts and modalities for all students, regardless of students' IEP status.

This is the only evaluation of the WG program that compares effectiveness of this program for students with and without IEPs. Given that WG is a freely available program that is being used in districts and schools nationally, it is essential that school leaders in urban districts with limited resources have a clear understanding how, and to what degree, this program can support their students with IEPs. The results here replicate the finding that students who participate in the program learn target words better than students in the comparison schools. We did not find that students with IEPs benefited more from program participation than students without them; however, we did find that the benefits that they experienced were sustained. In some respects, these findings are similar to our subgroup analysis of language-minority learners (Lawrence, Capotosto, Branum-Martin, White, & Snow, 2012). In that study, we found that the gains made by English-proficient language minority learners in the treatment schools were robust and sustained. One of the primary differences between that study and this one is that the overall learning trajectories of students in the sample did not vary as much by language status as they did by IEP status. The sobering reality of this analysis is that although students with IEPs seem to benefit from participation in WG, these one-shot differences are not enough to overcome the fanning of learning trajectories that results from underlying differences in vocabulary development across the calendar year. Ongoing studies of the program feature a wider range of measurement types that we hope will help us understand the steps that struggling readers need to take on the way to securing a rich and stable knowledge of academic words.

In this study, we also examined differences in students' vocabulary-learning trajectories associated with baseline achievement (measured by the state-mandated English language arts test). We found that although there were large differences in vocabulary at baseline associated with differences in achievement, differences on the MCAS did not predict subsequent underlying learning trajectories. Interestingly, differences in MCAS achievement were associated with how well students responded to the WG intervention. Although the main effect of the intervention was much greater than those effects related to student baseline performance, the differences in response were significant, whereas there was no such treatment by IEP status interaction.

There were several important limitations to this study. The first is that these results may only apply to schools with large numbers of students from low-SES homes; we have every reason to think that learning trajectories and summer setback, especially, will be different on average in wealthier school settings. Second, this study relies on district reports on students with IEPs, and we do not know how or when students were identified. Furthermore, the sample size was not large enough to investigate differences between students who had different designations under their IEPs. Future studies should include such analysis if possible, as well as systematic observations comparing students in content-area classes and substantially separate classrooms, to understand how the WG materials are being adapted by teachers to meet the needs of a range of students.

We were also concerned that the effect of MCAS on vocabulary may be different for students in special and general education programs, so we split the sample by median MCAS and fit the same model for high and low MCAS. The results were the same for the WG and IEP effects. We also ran the analyses separately for students with and without IEPs and found similar effects for WG. Although we acknowledge that the groups differ on covariates such as MCAS, our supplemental analyses suggest that these differences are not large. We do not have systematic observation data of instruction in either the treatment or comparison schools.

Other limitations include the fact that this study only included a small number of schools, and that the schools were not well matched; difference at baseline might have also contributed to differences in improvement as lower performing schools had more room to make vocabulary gains. Last, pretests were administered 2 to 3 months later in comparison schools, so in our primary efficacy study, we calculated main effects and also improvement per month (for full details, see Snow et al., 2009). Although these problems likely inflate main effects, they do not influence differential impacts or the longitudinal follow-up analysis presented here.

Despite these limitations, these results demonstrate that, for these students from low-SES homes, differences in general literacy achievement predict large differences in baseline academic vocabulary, but do not predict subsequent growth. Instead, IEP status does. Students with higher baseline literacy ability appear more able to learn from the WG program, although all participants made long-term improvements in their knowledge of high-leverage academic vocabulary words on average, and these improvements were sustained even a year after instruction. We take these findings as cautious support for the value of a cross-content discussion-based academic intervention for urban middle school students, and for the fact that this curricular intervention can support students with IEPs as well as student with low literacy skills.

REFERENCES

Armstrong v. Kline, 476 F. Supp. 583 (E.D. Pa. 1979).

August, D., Branum-Martin, L., Cardenas-Hagan, E., & Francis, D. J. (2009). The impact of an instructional intervention on the science and language learning of middle grade English language learners. *Journal of Research on Educational Effectiveness*, 2(4), 345–376. doi:10.1080/19345740903217623

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology*, *34*(6), 1373–1399. doi:10.1037/0012-1649.34.6.1373

Bhattacharya, A., & Ehri, L. (2004). Graphosyllabic analysis helps adolescent struggling readers read and spell words. *Journal of Learning Disabilities*, *37*(4), 331–348. doi:10.1177/00222194040370040501

Bos, C., & Anders, P. (1990). Effects of interactive vocabulary instruction on the vocabulary learning and reading comprehension of junior-high learning-disabled students. *Learning Disability Quarterly*, *13*(1), 31–42. doi:10.2307/1510390

Bryant, D., Goodwin, M., Bryant, B., & Higgins, K. (2003). Vocabulary instruction for students with learning disabilities: A review of the research. *Learning Disability Quarterly*, *26*(2), 117–129. doi:10.2307/1593594

Bulgren, J., Schumaker, J., & Deshler, D. (1988). Effectiveness of a concept-teaching routine in enhancing the performance of LD students in secondary-level mainstream classes. *Learning Disability Quarterly*, *11*(1), 3–17. doi:10.2307/1511034

Burns, M., & Boice, C. (2009). Comparison of the relationship between words retained and intelligence for three instructional strategies among students with below-average IQ. *School Psychology Review*, *38*(2), 284–292.

Cain, K., Oakhill, J., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of Child Language*, *30*(3), 681–694. doi:10.1017/s0305000903005713

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *96*(4), 671–681.

Ceci, S., & Papierno, P. (2005). The rhetoric and reality of gap closing: When the "have-nots" gain but the "haves" gain even more. *American Psychologist*, *60*(2), 149–160. doi:10.1037/0003-066x.60.2.149

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, *66*(3), 227–268. doi:10.3102/00346543066003227

Coxhead, A. (2000). A new academic word list. *Tesol Quarterly*, *34*(2), 213–238. doi:10.2307/3587951

Downing, J., & Peckham-Hardin, K. (2007). Inclusive education: What makes it a good education for students with moderate to severe disabilities? *Research and Practice for Persons with Severe Disabilities*, *32*(1), 16–30. doi:10.2511/rpsd.32.1.16

Ebbers, S., & Denton, C. (2008). A root awakening: Vocabulary instruction for older students with reading difficulties. *Learning Disabilities Research & Practice*, *23*(2), 90–102. doi:10.1111/j.1540-5826.2008.00267.x

Entwisle, D., Alexander, K., & Olson, L. (2000). Summer learning and home environment. In R. D. Kahlenberg (Ed.), *A notion at risk* (pp. 9–30). New York, NY: Century Foundation Press.

Folse, K. (2006). The effect of type of written exercise on L2 vocabulary retention. *Tesol Quarterly*, *40*(2), 273–293. doi:10.2307/40264523

Fore III, C., Boon, R., & Lowrie, K. (2007). Vocabulary instruction for middle school students with learning disabilities: A comparison of two instructional models. *Learning Disabilities: A Contemporary Journal*, *5*(2), 49–73.

Fukkink, R., Blok, H., & de Glopper, K. (2001). Deriving word meaning from written context: A multicomponential skill. *Language Learning*, *51*(3), 477–496. doi:10.1111/0023-8333.00162

Fukkink, R., & de Glopper, K. (1998). Effects of instruction in deriving word meaning from contexts: A meta-analysis. *Review of Educational Research*, *68*(4), 450–469. doi:10.3102/00346543068004450

Heyns, B. (1978). *Summer learning and the effects of schooling*. New York, NY: Academic Press.

Hock, M., Brasseur, I., Deshler, D., Catts, H., Marquis, J., Mark, C., & Stibling, J. (2009). What is the reading component skill profile of adolescent struggling readers in urban schools? *Learning Disability Quarterly*, *32*(1), 21–38.

Hughes, T., & Fredrick, L. (2006). Teaching vocabulary with students with learning disabilities using classwide peer tutoring and constant time delay. *Journal of Behavioral Education*, *15*(1), 1–23. doi:10.1007/s10864-005-9003-5

Kim, J. (2004). Summer reading and the ethnic achievement gap. *Journal of Education for Students Placed At Risk*, *9*(2), 169–188. doi:10.1207/s15327671espr0902_5

Kitajima, R. (2001). The effect of instructional conditions on students' vocabulary retention. *Foreign Language Annals*, *34*(5), 470–482. doi:10.1111/j.1944-9720.2001.tb02086.x

Lawrence, J. F. (2009). Summer reading: Predicting adolescent word learning from aptitude, time spent reading, and text type. *Reading Psychology*, 30(5), 445–465. doi:10.1080/02702710802412008

Lawrence, J. F., Capotosto, L., Branum-Martin, L., White, C., & Snow, C. E. (2012). Language proficiency, home-language status, and English vocabulary development: A longitudinal follow-up of the Word Generation program. *Bilingualism: Language and Cognition*, *15*(03), 437–451. doi:10.1017/s1366728911000393

Lovelace, S., & Stewart, S. (2009). Effects of robust vocabulary instruction and multicultural text on the development of word knowledge among African American children. *American Journal of Speech-Language Pathology*, *18*(2), 168–179. doi:10.1044/1058-0360(2008/08-0023)

Mastropieri, M., Scruggs, T., Levin, J., Gaffney, J., & McLoone, B. (1985). Mnemonic vocabulary instruction for learning disabled students. *Learning Disability Quarterly*, *8*(1), 57–63. doi:10.2307/1510908

McDonnell, J., Johnson, J., Polychronis, S., Riesen, T., Jameson, M., & Kercher, K. (2006). Comparison of one-to-one embedded instruction in general education classes with small group instruction in special education classes. *Education and Training in Developmental Disabilities*, *41*(2), 125–138. doi:10.1353/etc.2007.0001

McKeown, M. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly*, *20*(4), 482–496. doi:10.2307/747855

McKeown, M., Beck, I., Omanson, R., & Perfetti, C. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Reading Behavior*, *15*(1), 3–18. doi:10.1080/10862968309547474

McLeskey, J., & Henry, D. (1999). Inclusion: What progress is being made across states? *Teaching Exceptional Children*, *31*(5), 56–62.

McLeskey, J., Henry, D., & Hodges, D. (1999). Inclusion: What progress is being made across disability categories? *Teaching Exceptional Children*, *31*(3), 60–64.

Min, H. (2008). EFL vocabulary acquisition and retention: Reading plus vocabulary enhancement activities and narrow reading. *Language Learning*, *58*(1), 73–115. doi:10.1111/j.1467-9922.2007.00435.x

Morgan, P., Farkas, G., & Hibel, J. (2008). Matthew effects for whom? *Learning Disability Quarterly*, *31*(4), 187–198.

Mraz, M., & Rasinski, T. (2007). Summer reading loss. *Reading Teacher*, *60*(8), 784–789. doi:10.1598/rt.60.8.9

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

National Center for Education Statistics. (2013). *The nation's report card: A first look: 2013 mathematics and reading trial urban district assessment* (NCES 2014-466). Washington, DC: US Department of Education, Institute of Education Sciences.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: US Government Printing Office.

Penno, J., Wilkinson, I., & Moore, D. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology*, *94*(1), 23–33. doi:10.1037/0022-0663.94.1.23

Ricketts, J., Bishop, D., & Nation, K. (2008). Investigating orthographic and semantic aspects of vocabulary learning in poor comprehenders. *Journal of Research in Reading*, *31*(1), 117–135. doi:10.1111/j.1467-9817.2007.00365.x

Roberts, G., Torgesen, J., Boardman, A., & Scammacca, N. (2008). Evidence-based strategies for reading instruction of older students with learning disabilities. *Learning Disabilities Research & Practice*, *23*(2), 63–69. doi:10.1111/j.1540-5826.2008.00264.x

Ross, P. (1974). Regression study—Now the long hot summer. *Reading Teacher*, *28*(1), 28–30.

Rott, S. (2007). The effect of frequency of input-enhancements on vocabulary learning and text comprehension. *Language Learning*, *57*(2), 165–199. doi:10.1111/j.1467-9922.2007.00406.x

Seifert, K., & Espin, C. (2012). Improving reading of science text for secondary students with learning disabilities effects of text reading, vocabulary learning, and combined approaches to instruction. *Learning Disability Quarterly*, *35*(4), 236–247. doi:10.1177/0731948712444275

Shaywitz, B. A., Holford, T. R., Holahan, J. M., Fletcher, J. M., Stuebing, K. K., Francis, D. J., & Shaywitz, S. E. (1995). A Matthew effect for IQ but not for reading: Results from a longitudinal study. *Reading Research Quarterly*, *30*(4), 894–906. doi:10.2307/748203

Singer, J., & Willett, J. (2003). *Applied longitudinal data analysis: Modeling change and even occurrence*. New York, NY: Oxford University Press.

Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, *2*(4), 325–344. doi:10.1080/19345740903167042

Snow, C., Porche, M. V., Tabors, P., & Harris, S. (2007). *Is literacy enough?: Pathways to academic success for adolescents*. Baltimore, MD: Paul H. Brookes.

Stahl, S., & Fairbanks, M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, *56*(1), 72–110. doi:10.3102/00346543056001072

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–407. doi:10.1598/rrq.21.4.1

Strategic Educational Research Partnership. (2009). *Word generation*. Retrieved from http://www.wordgeneration.org

Swanborn, M., & de Glopper, K. (1999). Incidental vocabulary learning while reading: A meta-analysis. *Review of Educational Research*, *69*(3), 261–285. doi:10.3102/00346543069003261

Townsend, D., & Collins, P. (2009). Academic vocabulary and middle school English learners: An intervention study. *Reading and Writing*, *22*(9), 993–1019. doi:10.1007/s11145-008-9141-y

van Daalen-Kapteijns, M., Elshout-Mohr, M., & de Glopper, K. (2001). Deriving the meaning of unknown words from multiple contexts. *Language Learning*, *51*(1), 145–181. doi:10.1111/0023-8333.00150

Vaughn, S., Martinez, L., Linan-Thompson, S., Reutebuch, C., Carlson, C., & Francis, D. (2009). Enhancing social studies vocabulary and comprehension for seventh-grade English language learners: Findings from two experimental studies. *Journal of Research on Educational Effectiveness*, *2*(4), 297–324. doi:10.1080/19345740903167018

Wanzek, J., Vaughn, S., Roberts, G., & Fletcher, J. M. (2011). Efficacy of a reading intervention for middle school students with learning disabilities. *Exceptional Children*, *78*(1), 73–87.

Winchester, K., Darch, C., Eaves, R., Shippen, M., Ern, G., & Bell, B. (2009). An evaluation of two methods for teaching United States history to students with learning disabilities. *Journal of Direct Instruction*, *9*(1), 17.

Zhang, Z., & Schumm, J. (2000). Exploring effects of the keyword method on limited English proficient students' vocabulary recall and comprehension. *Literacy Research and Instruction*, *39*(3), 202–221. doi:10.1080/19388070009558322