UC Irvine
UC Irvine Previously Published Works

Title

Generating knowledge of academic language among urban middle school students

Permalink

https://escholarship.org/uc/item/0pp2m2z1

Journal

Journal of Research on Educational Effectiveness, 2(4)

Authors

Snow, Catherine E
Lawrence, Joshua Fahey
White, Claire

Publication Date

2009-10-01

Copyright Information

Peer reviewed

Routledge
Taylor & Francis Group

Generating Knowledge of Academic Language Among Urban Middle School Students

Catherine E. Snow, Joshua F. Lawrence, and Claire White
Harvard Graduate School of Education and Strategic Education Research Partnership,
Cambridge, Massachusetts, USA

Abstract: A quasi-experimental study of a novel, cross-content area vocabulary intervention program called *Word Generation* showed significantly greater growth among 6th- to 8th-grade students in schools implementing the program than in comparison schools, on a curriculum-specific test. Furthermore, the language-minority students in the treatment, but not the comparison, schools showed greater growth than the English-only students. Improvement on the curriculum-specific test predicted performance on the Massachusetts Comprehensive Assessment System (MCAS) English language arts assessment, again only for students in the treatment schools. Recognizing the need to implement a more rigorous experimental study of this program, nonetheless we conclude that participation in the intervention, with its focus on deep reading, comprehension of current-events topics, productive classroom discussion, developing arguments, and producing persuasive essays, was a plausible contributor to student performance on the MCAS.

Keywords: Vocabulary, academic language, discussion, reading comprehension, language minority students

The reading comprehension of post–primary grade students, in particular those attending urban schools, is a matter of recurrent concern. Performance of eighth-graders on the National Assessment of Educational Progress reading assessment, for example, shows that 74% of all students perform at or above the basic (grade-appropriate) level, whereas only 60% of students in the large central urban districts perform at that level. Ethnic and language minority students in these districts perform even less well than White and native English-speaking students. Although the specific reading challenges faced by students performing below basic level are no doubt heterogeneous, a ubiquitous issue mentioned by their teachers and confirmed by assessment is their limited vocabularies. It is not surprising that the many language minority students in urban districts show gaps in English vocabulary, but even native English speakers may fail to

Address correspondence to Catherine E. Snow, 7 Appian Way, Larsen Hall Room 307, Harvard University, Cambridge, MA 02138, USA. E-mail: snowcat@gse.harvard.edu

develop rich vocabularies if they have a history of low reading ability, limited comprehension, and low investment of time in reading, because much sophisticated vocabulary is acquired through reading (Anderson, Wilson, & Fielding, 1988; Stanovich, 1986).

In response to administrators' and teachers' worries about the vocabulary skills of students in the Boston Public Schools (BPS), an intervention called *Word Generation* was designed to promote the learning of all-purpose academic vocabulary among sixth- to eighth-grade students. Word Generation was developed under the auspices of the Strategic Education Research Partnership (SERP), as part of the BPS–SERP field site activities (www.serpinstitute.org). Word Generation adheres to research-based principles of vocabulary learning and incorporates opportunities for students to use the five words taught each week in classroom discussion, debate, and writing by embedding those words in brief texts that present controversies of high interest to (pre)adolescents. Thus, in addition to teaching vocabulary, the program provides opportunities for students to develop and practice oral academic language skills, argumentation strategies, and writing skills while learning about issues of current public interest (e.g., steroid use among athletes, legalization of euthanasia, censorship of libraries and of popular music, funding for stem cell research).

In this article we report on an analysis of the impact of Word Generation on students' learning of words taught and the relationship of their performance on curriculum-linked assessments of Word Generation vocabulary to their performance on the English Language Arts assessment administered as part of the state accountability system. In addition, in light of the greater risk for poor literacy performance of students from non-English-speaking homes, we analyze the differential effect of the program on language minority (LM) and English-only (EO) students.

PRINCIPLES UNDERLYING WORD GENERATION

A first principle of effective vocabulary instruction is to teach words students need to know—those that are crucial to their comprehension of their assigned texts, for example. Middle school students are expected to read content-area texts that contain many technical, discipline-specific words as well as many "all-purpose academic words." This latter category of words is less likely to be explicitly taught, in particular by math, science, and social studies teachers who concentrate their instructional time and effort on the words of their respective disciplines. Yet students who do not know the all-purpose academic words struggle with comprehension of texts containing them (Anderson & Freebody, 1981), and thus struggle as well to learn the disciplinary concepts and words that are being taught. For these reasons, a basic Word Generation design principle was to focus on all-purpose academic words but to enhance content area teachers' willingness to teach them.

Additional Word Generation features were designed to implement what we know about effective vocabulary teaching. Fortunately, the field of vocabulary instruction has been well researched. Dozens of small-scale experimental studies provide evidence about instructional factors that promote successful vocabulary learning (Beck, McKeown, & Kucan, 2002; Beck, Perfetti, & McKeown, 1982; Graves, 2006; McKeown, Beck, Omanson, & Perfetti, 1983; McKeown, Beck, Omanson, & Pople, 1985; National Institute of Child Health and Human Development, 2000; Stahl & Fairbanks, 1986; Stahl & Nagy, 2006). Those factors include the following:

- Encountering the target word in semantically rich contexts within motivating texts, rather than in a list of words.
- Recurrent exposure to the word, in varied contexts.
- Opportunities to use the word orally and in writing.
- Explicit instruction in word meaning.
- Explicit instruction in word learning strategies, including morphological analysis, cognate use, and polysemy.

Word Generation was designed to provide instruction that embodied these factors. Thus, the target words were presented in brief but engaging texts, and teachers were given guidance about how to read those texts with their students so as to ensure comprehension. Then explicit instruction in the word meanings was provided. The words were also embedded in activities to be carried out by math, social studies, and science teachers, to ensure both recurrent exposure and opportunities to learn about subject-specific meanings (polysemy) and contexts of use. The daily activities were designed to promote opportunity for oral discussion and debate on 4 days of the week, and writing on the 5th.

DESIGN OF THE INTERVENTION

Word Generation is a 24-week-long sequence of topics of current interest, each associated with five all-purpose academic words, and prescribed activities related to math, science, and social studies (see www.wordgeneration.org for sample materials). The basic sequence of Word Generation activities was the following: On Monday a brief text in which the five target words were embedded was read by the students and teacher together, then discussed using guiding comprehension questions; this text presented arguments on both sides of some difficult controversy or dilemma. Then the five target words were highlighted and provided with student-friendly context-related definitions. This activity typically occurred in the English Language Arts classroom. On Tuesday, Wednesday, and Thursday, in an order determined by each school, the math, social studies, and science teachers respectively implemented activities provided for them, each of which embedded the same five target words. The

math teacher assigned one or two problems related in content to the dilemma of the week; the format of these problems was modeled on the state math assessment. Math teachers then discussed the content as well as the math procedures. The science teacher presented a new text that focused on science content related to the dilemma of the week; students filled in target words left blank in the text before the class discussed the text. The social studies teacher organized a debate about the dilemma of the week in one of several possible formats (fishbowl, pairs, whole class, four corners, etc.). On Friday, the students were asked to write a "taking a stand" essay about the dilemma.

Various aspects of the Word Generation design respond to the local conditions in the district for which it was originally developed. Most middle-grade students in the BPS attend separate middle schools serving sixth to eighth graders, though about 25% are in K–8 schools. In the middle schools, content area instruction is departmentalized, and teacher planning time built into the school schedule typically occurs within departments, limiting the opportunities for teachers to share information about student progress or curricular emphases across those departmental boundaries. Transcending the boundaries to recruit participation by all the teachers in vocabulary teaching was one goal of the Word Generation design.

In addition, the extensively articulated state and district curriculum standards, as well as district pacing guides for math, science, and social studies, limited the classroom time available for focusing on vocabulary or on topics not explicitly included in the standards. Thus, to secure collaboration from the District leadership and the teachers, we agreed to design activities that could be completed in 15 min per day (thus taking only 15 min *per week* from math, science, or social studies). Furthermore, each school implementing the program had considerable leeway to decide on scheduling (which group of teachers was responsible for which day[s] of the week) and on extent of use. For example, one school excluded Structured English Immersion students during the 1st year of implmentation but included them subsequently. The five schools reported on here used the program with all students in Grades 6 to 8.

effects on students participating in the program, acknowledging that the basis for strong causal inference about program impact is limited in this quasi-experimental study. We posed three research questions:

1. Did students in the schools implementing Word Generation learn the words taught?
2. Was there a differential effect of Word Generation on LM and EO students?
3. Did the learning promoted by Word Generation relate to performance on external achievement measures?

METHODS

This is a quasi-experimental study in which academic word-learning by students in five schools implementing the Word Generation program was compared to academic word-learning by students in three schools within the same system that did not choose to implement the program. Because the implementing schools were those that volunteered for the program, selection effects must be taken into account in interpreting the findings. The data reported in Table 1 suggest that the comparison schools were performing better than the treatment schools at the start of the current study and that impression was confirmed by disparities in performance on the curriculum-specific pretest (see Figure 1). The design of the intervention precluded assignment to treatment at the individual or classroom level.
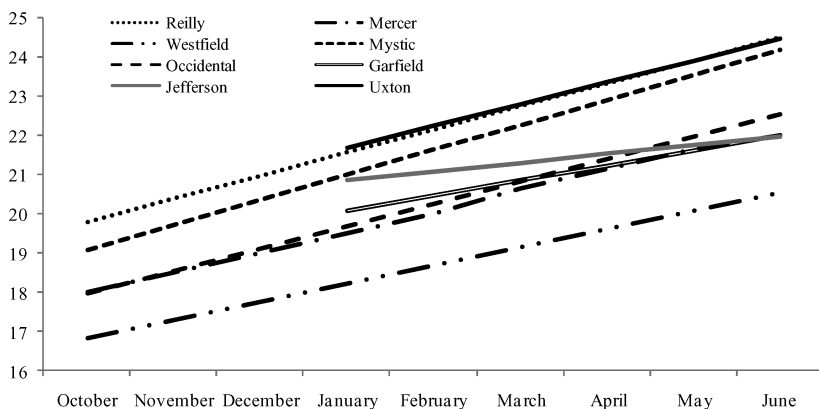

Figure 1. Improvement on vocabulary measure by school.

Table 1. Information about the participating schools

School	% of Students From LM Homes	% of Students With IEPs	% of Students Receiving Free of Reduced Lunch	Level of Participation in PD	ELA MCAS 2007				ELA MCAS 2008				Score
					W*	NI	P	A	W	NI	P	A	
Treatment schools													
Reilly	31.9	26.3	82.7	High	10.7	33.7	52.3	4.0	6.3	33.3	56.0	4.7	242.4 (12.4)
Mercer	43.9	25.4	87.4	Low	16.3	40.3	43.0	1.0	18.7	35.3	42.3	4.0	238.2 (14.2)
Westfield	48.8	27.1	78.8	Medium	27.0	42.0	30.0	0.7	27.7	45.3	27.3	0.0	233.7 (11.5)
Mystic	69.7	16.1	89.7	Medium	2.3	30.3	63.3	4.0	5.7	30.0	58.7	5.7	241.1 (13.0)
Occidental	51.7	32.6	90.7	High	35.3	38.7	26.0	0.0	42.0	33.3	24.3	0.3	234.0 (13.5)
Comparison schools													
Garfield	54.0	31.7	81.0	N/A	3.0	26.3	68.3	2.0	5.3	30.0	61.3	3.7	240.3 (12.1)
Jefferson	35.8	30.7	78.1	N/A	4.0	39.0	51.0	7.0	9.0	45.5	37.0	8.0	242.2 (16.2)
Uxton	61.4	19.7	89.8	N/A	21.3	39.0	37.0	3.0	26.3	32.7	39.0	2.3	239.4 (10.1)
Average scores													
Treatment	49.2	25.5	85.8		18.3	37.0	42.9	1.9	20.1	35.5	41.7	2.9	239.3 (13.5)
Comparison	50.4	27.4	83.0		9.4	34.8	52.1	4.0	13.6	36.1	45.8	4.7	240.3 (12.3)

Note. Only P and A are considered passing. ELA = English Language Arts; MCAS = Massachusetts Comprehensive Assessment System; LM = language minority; IEP = Individualized Education Plan; PD = professional development; W = warning, NI = needs improvement, P = proficient, A = advanced.

School Demographics

Word Generation was implemented in five schools during the 2007–2008 academic year, three middle schools and two K–8 schools in which only the sixth to eighth grades used the program. Two schools, the Reilley and the Westfield, were completing their 2nd year of implementation in 2007–2008, whereas the Mystic, Occidental, and Mercer Schools launched Word Generation in fall 2007 (pseudonyms are used for all schools). Demographics of the Word Generation and comparison schools reflect BPS more broadly, with a high incidence of poverty (ranging from a low of 79% to a high of 91% students receiving free or reduced-price lunch; see Table 1). BPS is characterized in general by rather high levels of special education designation, and all the schools shared this feature (between 16% and 33% of students with Individualized Education Programs). A very high proportion of students at these schools come from second-language homes, with percentages ranging from 32% to 70% across the schools. Four of the treatment schools offered Sheltered English Immersion services to their limited English proficient students; all students enrolled in these sheltered classrooms (who represented between 6% and 26% of their school populations) received the Word Generation curriculum, albeit with modifications such as extended time and translation of key concepts.

The comparison schools looked somewhat less disadvantaged as a group than the intervention schools (see Table 1), and their average scores on the state accountability assessment at the start of this study were higher (mean of 45% failing in the comparison schools, compared to 56% in the treatment schools). This is not surprising; the schools volunteered to participate in the intervention, and those with lower scores were more likely to show an interest.

Professional Development (PD) and Implementation

In addition to varying on demographic profiles, the treatment schools varied in the degree to which they utilized the PD offered to support implementation of Word Generation, and in the eagerness with which the teachers across the building embraced the program. The full PD package available to the schools involved school leader participation in an implementation planning session, a schoolwide training for all teachers prior to starting the program, two to three sessions of implementation support throughout the year, and regularly scheduled school-based sessions of collaborative planning and preparation. Specifically,

- Key instructional leaders in newly implementing schools were invited to a Word Generation summer institute, where they could organize testing and implementation schedules with the Word Generation research team.

- Prior to launching the intervention, a 4-hr PD session is recommended. These hours are devoted to introducing the yearlong intervention, discussing relevant research on vocabulary teaching and learning, viewing videotapes of exemplary implementation of the program by other BPS practitioners, and hands-on practice with program materials.
- Ongoing PD (two to three more school-based sessions) is also recommended, and most schools establish several dates throughout the academic year when feedback and further PD can be delivered.
- Effective implementation of Word Generation is contingent upon a set of shared understandings and commitments at the school level. Regular cluster, grade, and/or content level meetings for team building, updating implementation schedules and calendars, and other organizational features are vital for ensuring cross-content communication and maintaining effective program implementation.

As can be seen from Table 1, the five implementing schools participated in these various activities to varying degrees because of difficulties scheduling and organizing the required meetings. For example, the Mercer received only one brief PD session, whereas the Occidental participated in the summer institute, received 8 hr of PD prior to launch, and engaged in biweekly cross-grade school-site sessions devoted to feedback on and previewing of the materials and activities, with support from the Word Generation team at several of those sessions.

Student Participants

Both pre- and posttest data were available on 697 sixth-, seventh-, and eighth-grade students in five treatment schools and 319 in three comparison schools. All students in the treatment schools received the intervention; those represented in this data set had completed usable test forms at both pre- and posttest. There were 349 girls and 348 boys in the treatment schools, and 162 girls and 157 boys in comparison schools. Of these, 438 were classified as LM (parents reported preferring to receive materials in a language other than English), 287 in treatment schools and 151 in comparison schools. As can be seen from Table 1, the vast majority of students in both treatment and comparison schools were low income.

Measures

The efficacy of the intervention was assessed using a 48-item multiple-choice test that randomly sampled two of the five words taught each week. The vocabulary assessment was not completed by all students in the time available.

Because items at the end of the assessment had particularly low rates of completion, we dropped the last 4 items from our analysis of both pre- and posttest. The reliability of the test with the 40 items that remained was acceptable (Cronbach's $\alpha = .876$).

This instrument was administered to students in all the treatment schools in October 2007, before the introduction of Word Generation materials. Because of difficulty recruiting the comparison schools, the pretest was not administered there until January. The posttest (identical to the pretest except for the order of items) was administered in all the schools in late May. Because of the unfortunate disparity in interval between pre- and posttesting in the two groups of schools, we present analyses in terms of words learned per month as well as total words learned.

In addition to this curriculum-based assessment, we had access for most of the students to scores on the Massachusetts Comprehensive Assessment System (MCAS) English Language Arts (ELA) scores for spring 2008. In addition, we had Group Reading and Diagnostic Evaluation (GRADE; Williams, 2000) for both spring and fall for a selection of students in all comparison ($n = 133$) and treatment ($n = 256$) schools (Table 2). These scores were provided by the district for all the students for whom data were available. The decision to adminster the assessment was made at the school and classroom level. Thus, although these data are far from complete, we have no reason to think that there was a particular sampling bias across the schools.

Analysis

Our general approach to the analysis was to use regression models to predict posttest scores, controlling for pretest scores as well as gender and subsequently introducing variables of interest (treatment condition, language status) as well as their interactions. We recognize that this analytic strategy ignores the nesting of students within schools, and thus planned follow-up analyses comparing treatment and comparison schools as well as within one, large treatment school to further explore the initial findings.

RESULTS

Descriptive statistics suggested that boys did better than girls on the pre- and posttest, that the pretest scores of students in the comparison schools were higher than those in treatment schools, and that LM students performed less well than EO students across all the schools (Table 3). These results show that students in the Word Generation program learned approximately the number of words that differentiated eighth from sixth graders on the pretest—in other

Table 2. Spring and fall group reading and diagnostic evaluation score for participating schools

School	Fall	Spring	Valid N
Treatment schools			
Reilly	463.72	476.61	54
	(33.27)	(35.65)	
Mercer	460.07	475.27	116
	(25.31)	(25.56)	
Westfield	452.47	457.76	34
	(27.13)	(34.55)	
Mystic	456.57	466.77	35
	(18.89)	(20.86)	
Occidental	446.42	458.12	17
	(22.45)	(27.00)	
Average	458.45	470.81	256
	(26.75)	(29.61)	
Comparison schools			
Garfield	456.13	468.07	14
	(15.76)	(21.28)	
Jefferson	471.42	482.50	60
	(34.69)	(35.53)	
Uxton	464.75	479.05	59
	(17.81)	(24.96)	
Average	466.65	479.81	133
	(26.88)	(29.14)	

words, participation in 20 to 22 weeks of the curriculum was equivalent to 2 years of incidental learning.

Unfortunately, the relative improvements in the Word Generation schools will be exaggerated by the differences in timing of the pretest. Table 4 presents both the total pre- to posttest improvement and the improvement divided by the number of months between pre- and posttests (8 months for treatment schools, 5 months for comparison schools). The results shown in Table 4 demonstrate that Word Generation schools outperformed the comparison schools when the amount of time between tests is taken into consideration (see Figure 1).

Regression analysis was used to determine if participation in Word Generation predicted improved vocabulary outcomes, controlling for the pretest. Gender was a significant predictor of word learning ($\beta = -0.052$, $p < .007$), as was treatment ($\beta = 0.166$, $p < .001$; see Table 5). Language status (LM vs. EO) was not a significant predictor, but the interaction of treatment and language status was at the margin of significance ($p = .055$), and including the interaction improved the overall model. Of interest, student pretest vocabulary did not interact with treatment in predicting posttest scores.

Table 3. Pre- and posttest scores for treatment and comparison students by grade, gender, and language status

Treatment	Grade Level	All Students		Gender				Language Status			
				Male		Female		LM		EO	
		Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Word generation students	6th	17.0	22.0	16.7	21.6	17.4	22.4	17.1	22.6	17.0	21.6
	7th	18.6	23.1	19.6	24.2	17.7	22.1	18.5	23.2	18.7	23.1
	8th	20.6	24.4	22.0	25.9	19.1	22.7	19.9	24.3	21.2	24.4
Comparison students	6th	19.9	22.0	20.2	22.9	19.6	21.2	19.4	21.9	20.4	22.2
	7th	21.5	24.0	22.2	25.1	20.9	22.9	21.2	23.4	21.8	24.4
	8th	23.1	25.2	22.9	26.5	23.2	24.1	22.6	24.0	23.6	26.7

Note. LM = language minority; EO = English-only.

Table 4. Improvement on vocabulary measure, improvement per month, and effect sizes for all schools by treatment

School	Pretest	Posttest	Improvement	Improvement/ Month	Effect Size
Treatment schools					
Reilly	19.79	24.51	4.72	0.59	0.56
	(6.54)	(6.77)	(4.08)		
Mercer	18.01	22.02	4.01	0.50	0.40
	(6.14)	(7.15)	(4.16)		
Westfield	16.85	20.55	3.70	0.46	0.33
	(6.29)	(7.39)	(5.54)		
Mystic	19.08	24.2	5.12	0.64	0.65
	(6.13)	(6.65)	(4.66)		
Occidental	17.98	22.56	4.58	0.57	0.53
	(6.36)	(7.2)	(4.32)		
Average	18.64	23.07	4.43	0.55	0.49
	(6.33)	(6.85)	(4.33)		
Comparison schools					
Garfield	20.07	22	1.93	0.39	
	(6.48)	(7.3)	(5.02)		
Jefferson	20.85	21.97	1.12	0.22	
	(7.7)	(8.06)	(3.27)		
Uxton	21.67	24.47	2.80	0.56	
	(5.62)	(5.92)	(4.64)		
Average	21.23	23.45	1.95	0.39	
	(6.38)	(6.85)	(4.33)		

We split the data set to investigate the home language variable more closely (Table 6). The first set of regressions used pretests and gender to predict posttest scores in the comparison schools ($r^2 = .62$) and Word Generation schools ($r^2 = .64$). In Word Generation schools LM status predicted improved

Table 5. Predicting vocabulary posttest scores for the whole sample with gender, treatment, and home language status with interaction term, controlling for vocabulary pretest

R^2	Coefficients	B	SE B	β	t Statistic	p
	(Constant)	5.165	0.589		8.764	<.001
(.61)	Pretest	0.870	0.021	0.801	41.031	<.001
(.61)	Gender	−0.723	0.269	−0.052	−2.684	.007
(.63)	Treatment	2.502	0.432	0.166	5.788	<.001
(.63)	EO	0.337	0.479	0.024	0.704	.482
.629	Treatment × EO	−1.116	0.581	−0.078	−1.921	.055

Note. EO = English-only.

Table 6. Predicting vocabulary posttest scores in comparison versus Word Generation schools, from gender, treatment, and language minority status with interaction term, controlling for vocabulary pretest

Treatment	R^2	Coefficients	*B*	*SE B*	*β*	*t* Statistic	*p*
Comparison		(Constant)	6.367	0.899		7.085	<.001
schools	.608	Pretest	0.832	0.037	0.774	22.252	<.001
	.620	Gender	−1.465	0.476	−0.107	−3.078	.002
	.620	EO	0.331	0.477	0.024	0.695	.488
Word		(Constant)	7.110	0.581		12.246	<.001
Generation	.634	Pretest	0.890	0.026	0.795	34.603	<.001
schools	.635	Gender	−0.367	0.326	−0.026	−1.125	.261
	.638	EO	−0.756	0.330	−0.053	−2.293	.022

Note. EO = English-only.

vocabulary ($β = -0.053$, $p = .022$), but it was not a significant predictor in comparison schools. The growth for LM and EO students in comparison versus Word Generation schools is plotted in Figure 2.
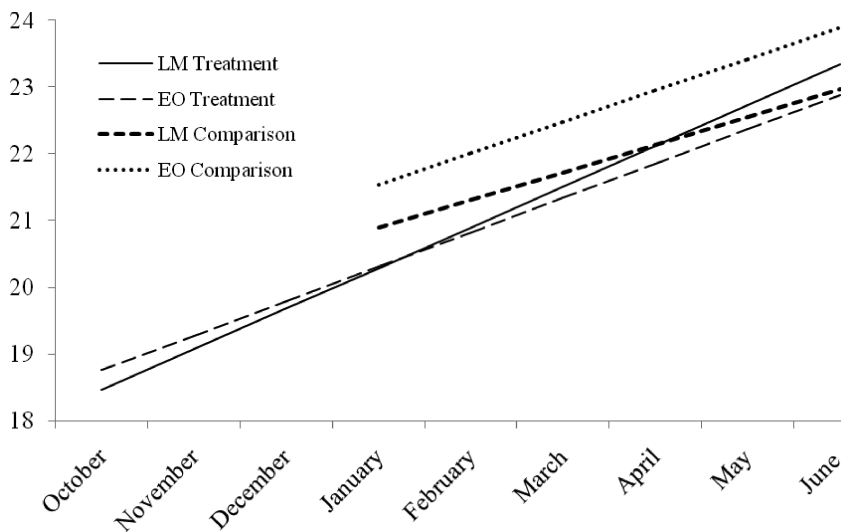

Figure 2. Vocabulary improvement for English-only (EO) and language minority (LM) students in Word Generation and comparison schools.

Predicting Standardized Test Scores

To determine whether participation in Word Generation had any relationship to performance on the MCAS, a regression model was fit with MCAS scores in April 2008 as the outcome, using gender, treatment status, pretest, and posttest scores as predictors (Table 7). We added an interaction term to see if posttest scores interacted with treatment in predicting MCAS scores (controlling for pretest scores). The interaction term was significant ($\beta = .21, p = .01$) and its inclusion improved the model.

We further explored the interaction between treatment and vocabulary improvement by splitting the data and refitting the models to data from the treatment and comparison school separately. The fitted model for comparison school data did not predict MCAS achievement ($R^2 = .41$) as well as the fitted model for the treatment school data ($R^2 = .49$). In the Word Generation schools student posttest scores ($\beta = 0.527, p < .001$) were much stronger predictors of MCAS achievement than pretest scores were ($\beta = 0.201, p < .001$), perhaps because the posttest scores captured not only target vocabulary knowledge at the end of the year but also level of student participation in the Word Generation program (Table 8).

Unfortunately, these analyses do not control for baseline reading achievement data, which were available only for a subset of students in our sample ($n = 389$). For that subgroup, we used fall standardized reading comprehension scores (on the GRADE) as a covariate to determine if the relation between improved vocabulary and MCAS persisted even when controlling for overall reading levels. Results demonstrate both that the GRADE is a strong predictor of spring MCAS scores ($\beta = 0.750, p < .001$) and that the interaction between treatment and improvement persists in the model controlling for GRADE. Split file analysis demonstrated the familiar pattern, with vocabulary improvement predicting MCAS scores for student in the treatment schools ($\beta = 0.151, p < .001$) but not for students in the comparison schools (see Tables 9 and 10).

Table 7. Predicting massachusetts comprehensive assessment system from gender, treatment status, pre- and posttest, and Treatment × Posttest interaction term

R^2	Coefficients	B	$SE\ B$	β	t Statistic	p
	(Constant)	209.06	1.96		106.80	<.001
(.012)	Gender	4.96	0.59	0.19	8.38	<.001
(.614)	Treatment	−4.53	2.24	−0.16	−2.02	.04
(.629)	Pretest	0.50	0.08	0.25	6.62	<.001
(.629)	Posttest	0.77	0.09	0.41	8.12	<.001
(.631)	Posttest × Treat	0.23	0.09	0.21	2.48	.01

Table 8. Predicting Massachusetts Comprehensive Assessment System scores from vocabulary pre- and posttests in comparison and word generation schools

Treatment	R^2	Coefficients	B	SE B	β	t Statistic	p
Comparison schools	.408	(Constant)	212.129	1.994		106.376	<.001
		Pretest	0.673	0.134	0.349	5.043	<.001
		Posttest	0.590	0.124	0.328	4.742	<.001
Word Generation schools	.486	(Constant)	207.992	1.278		162.756	<.001
		Pretest	0.431	0.096	0.201	4.472	<.001
		Posttest	1.011	0.086	0.527	11.736	<.001

GRADE scores were also used to determine if better readers learned words more efficiently than less able readers. Results demonstrate that GRADE baseline scores did not predict word learning and that there was no significant interaction between treatment and baseline reading achievement as measured on the GRADE (Table 11).

School-Level Confound

One of the challenges in interpreting these data is that there are significant differences both in the size of the program impact in different Word Generation schools (ranging from 3.7 to 5.1 points improvement on average) and in the language demographics (percentage of LM students) of those schools, making it difficult to disentangle student language status from school treatment effects. To address this issue, we explored language status as a predictor of word learning within the largest school in our sample, a school that also has sizeable

Table 9. Predicting Massachusetts Comprehensive Assessment System scores from gender, treatment status, and vocabulary improvement with treatment by improvement interaction term, controlling for fall GRADE score

R^2	Coefficients	B	SE B	β	t Statistic	p
	(Constant)	51.076	8.431		6.058	<.001
(.752)	Fall GRADE	0.409	0.018	0.750	22.782	<.001
(.755)	Treatment	−0.555	1.319	−0.018	−0.421	.674
(.764)	Improvement	0.016	0.231	0.005	0.070	.944
.766	Improvement × Treatment	0.494	0.263	0.149	1.881	.061

Note. GRADE = Group Reading and Diagnostic Evaluation.

Table 10. Predicting Massachusetts Comprehensive Assessment System scores from gender and vocabulary improvement in Comparison and Word Generation schools, controlling for fall GRADE score

Treatment	R^2	Coefficients	B	SE B	β	t Statistic	p
Comparison schools[a]	(.688) .688	(Constant) Fall GRADE Improvement	76.321 0.355 0.016	15.271 0.033 0.242	0.688 0.004	4.998 10.865 0.065	<.001 <.001 .948
Word Generation schools[b]	(.784) .799	(Constant) Fall GRADE Improvement	38.234 0.436 0.493	9.698 0.021 0.122	0.767 0.151	3.943 20.511 4.039	<.001 <.001 <.001

Note. GRADE = Group Reading and Diagnostic Evaluation.
[a]$n = 134$. [b]$n = 265$.

numbers of both LM ($n = 110$) and EO ($n = 179$) students. Within this school, speaking a language other than English at home predicted better vocabulary posttest scores controlling for pretests ($\beta = 0.073$, $p = .034$). Thus, though school treatment effects complicate our understanding of how language status interacts with participation in Word Generation, there is evidence from at least one large school that the LM students are indeed benefitting differentially from Word Generation.

Table 11. Predicting word learning from GRADE scores

R^2	Coefficients	B	SE B	β	t Statistic	p
.034	(Constant)	1.459	3.861		0.378	.706
	Fall GRADE	0.006	0.008	0.034	0.669	.504
.293	(Constant)	−3.826	3.800		−1.007	.315
	Fall GRADE	0.013	0.008	0.078	1.600	.110
	Treatment	2.840	0.469	0.294	6.059	<.001
.298	(Constant)	2.282	6.594		0.346	.729
	Fall GRADE	0.000	0.014	−0.001	−0.009	.993
	Treatment	−6.215	8.005	−0.644	−0.776	.438
	GRADE × Treatment	0.020	0.017	0.931	1.133	.258

Note. GRADE = Group Reading and Diagnostic Evaluation.

DISCUSSION

The results of this initial trial of a novel approach to teaching academic language and vocabulary are promising. Students in schools implementing the program learned more of the targeted words than students in comparison schools, even though the latter group performed at a higher level at the start. Language minority students benefited more strongly than EO students, and improvement on the curriculum-specific assessment predicted performance on the state ELA assessment. Although the design of this study precludes making strong causal inferences, these preliminary results are encouraging. In particular, though the significant differences in the language demographics of different Word Generation schools makes it difficult to disentangle effect of student language status and school treatment effects, the LM–EO differences in word learning were replicated within one school. This analysis suggests that confounding effects of school-level effectiveness do not explain the faster word learning of LM students. Instead, we may need to contemplate the possibility that these students were benefiting from effective, engaging, vocabulary-focused pedagogy.

It is of interest to compare the effect size obtained with the Word Generation curriculum to that obtained in other vocabulary interventions. A similarly structured intervention, the Vocabulary Improvement Program (Carlo et al., 2004), obtained an effect size of .50. The Stahl and Fairbanks (1986) meta-analysis of vocabulary curricula reviewed studies with effect sizes ranging as high as 2 under short-term laboratory-teaching conditions and as low as 0 under more authentic educational conditions. Thus, although Word Generation is not just a vocabulary intervention, and by design did not try to teach large numbers of words, its impact on students compares well with that of other successful programs.

It is particularly encouraging that posttest scores on the Word Generation assessments strongly related to performance on the state accountability assessment. One might assume this reflects the coincidence that the words taught also occurred on the state test. However, this simple explanation is undermined by the absence of a similarly strong relationship in the treatment schools. Furthermore, although improvement in the Word Generation schools was significant, it was still modest—about 4 words out of 40 tested. That translates into only about 12 words out of the 120 taught, which can hardly by itself explain a lot of variance on a long and challenging ELA assessment. Rather, we think it likely that improvement on our curriculum-based assessment represents an index of exposure to the Word Generation curriculum—a curriculum that taught new content, deep reading and comprehension skills, discussion, argumentation, and writing. Because the Massachusetts test is a relatively challenging one (arguably the best aligned with the National Assessment of Educational Progress of all the state assessments; McBeath, Reyes, & Ehrlander, 2007), performance

on the MCAS is more likely to be related to those complex skills than to specific word knowledge.

A major design feature of Word Generation is the effort to organize instruction around engaging topics and issues. Of course, some of those topics were more successfully engaging than others, a factor that might be related to the success of the students in actually learning the words associated with them. Looking at the words that showed particularly strong gains (percentage of students knowing them at posttest minus percentage knowing them at pretest), and in particular strong relative improvement (ratio of improvement in Word Generation schools to improvement in comparison schools), may provide some hints about the topics that were most salient and engaging. For example, 2 of the 10 best learned words (*amnesty*, *exclude*) came from the topic "undocumented immigrants," one likely to be of particular interest to the LM students. The other topic that generated two high-ranking words was "paying students to do well in school" (*incentive*, *enable*). Words from topics directly related to adolescence—grades as a criterion for sports participation, children's health/obesity, academic tracking, funding for girls in sports, preventing bullying—were among the better learned, whereas the topic of compulsory voting was associated with the two least well-learned words (*enforce*, *apathy*). Of course, it is difficult to disentangle features of the topics from quality of the activities associated with them, the teacher interest in them, and the learnability of the words themselves. In future research we will collect data from students themselves about their interest in the array of topics presented, as a mechanism to inform us about what makes some topics engaging and to test to what degree topic explains word learning.

Limitations

This study has many limitations. Our attempt to carry out a well-designed quasi-experiment was somewhat compromised by the fact that one comparison school in particular was performing much better than any intervention school, reflecting that school's emphasis on vocabulary across the curriculum. The design was also threatened by the fact that we received access to the comparison schools only partway through the school year, so pretest schedules were disrupted. The implementation schools varied as well in their capacity to access and benefit from the professional development opportunities offered and in their capacity to collect pre- and posttest data from large numbers of students. Though we had GRADE data available as a covariable, it was missing for a large percentage of students in both intervention and comparison schools. We were constrained by the small number of schools and the varying numbers of students across schools, and thus could not fully address the nesting of students within schools in our data analysis.

We are addressing these limitations in our current and imminent work on Word Generation. We have extended the quasi-experimental study to another year, with more schools and tighter control over the data collection procedures. We will be mounting a proper experimental study of Word Generation's impact and will explore how to support its effective implementation at larger scale. We have collected videotapes both of professional development sessions and of teachers implementing various Word Generation lessons to serve as a support for professional development (http://www.wordgeneration.org) or a self-teaching tool. Considerable work also needs to be done to document quality of implementation, to improve available materials, and to extend the curricular topics and activities. Nonetheless, we are greatly encouraged by these preliminary data suggesting the effectiveness of Word Generation in general, and its particular its value for LM learners.

ACKNOWLEDGMENTS

REFERENCES

Anderson, R.C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews*. Newark, DE: International Reading Association.

Anderson, R., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly*, *23*, 285–303.

Beck, I., McKeown, M., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford.

Beck, I., Perfetti, C., & McKeown, M. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, *74*(4), 506–521.

Carlo, M.S., August, D., McLaughlin, B., Snow, C.E., Dressler, C., Lippman, D.N., Lively, T.J., & White, C.E. (2004). Closing the gap: Addressing the vocabulary needs of English language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, *39*(2), 188–206.

Graves, M. (2006). *The vocabulary book: Learning and instruction*. New York: Teacher's College Press.

McBeath, J., Reyes, M. E., & Ehrlander, M. F. (2007). *Education reform in the American states*. Charlotte, NC: IAP.

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.

McKeown, M., Beck, I., Omanson, R., & Pople, M. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly*, *20*(5), 522–535.

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: U.S. Government Printing Office.

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

Stahl, S., & Nagy, W. (2006). *Teaching word meanings*. Mahwah, NJ: Erlbaum.

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

Williams, K. T. (2000). *Group reading assessment and diagnostic evaluation*. Circle Pines, MN: American Guidance Service.