THE SENIOR URBAN EDUCATION RESEARCH FELLOWSHIP SERIES

Word Generation in Boston Public Schools: Natural History of a Literacy Intervention

Catherine E. Snow
Joshua F. Lawrence

Council of the Great City Schools

VOLUME III
SPRING 2011
THE COUNCIL OF THE GREAT CITY SCHOOLS

The Council of the Great City Schools thanks the Institute of Education Sciences (IES) for supporting the Senior Urban Education Research Fellowship Program.

The findings and conclusions presented herein are those of the authors and do not necessarily represent the views of the Council of the Great City Schools or IES.

The Senior Urban Education Research Fellowship Series

Volume III:

Word Generation in Boston Public Schools: Natural History of a Literacy Intervention

by Catherine E. Snow and Joshua F. Lawrence
Spring 2011

The Council of the Great City Schools is the only national organization exclusively representing the needs of urban public schools. Founded in 1956 and incorporated in 1961, the Council is located in Washington, D.C., where it works to promote urban education through legislation, research, media relations, instruction, management, technology, and other special projects.

TABLE OF CONTENTS

FIGURES AND TABLES

OVERVIEW

OVERVIEW

THE SENIOR URBAN EDUCATION RESEARCH FELLOWSHIP PROGRAM

Large urban public school districts play a significant role in the American education system. The largest 65 urban school systems in the country – comprising less than one half of one percent of the nearly seventeen thousand school districts that exist across the United States – educate about 14 percent of the nation's K-12 public school students, approximately a third of its African American students, a quarter of its Hispanic students, a third of its limited English proficient students, and about a quarter of its economically disadvantaged students.[1] Clearly, any attempt to improve achievement and to reduce racial and economic achievement gaps across the United States must involve these school districts as a major focus of action.

These school districts face a number of serious, systematic challenges. To better understand the problems in urban education and to develop more effective and sustainable solutions, urban districts need a program of rigorous scientific inquiry focusing on what works to improve academic outcomes in the urban context. Moreover, in order to produce such evidence and to move public education forward generally, the standards of evidence in education research must be raised in such a way as to bring questions regarding the effectiveness of educational interventions and strategies to the fore and to promote careful scrutiny and rigorous analysis of the causal inferences surrounding attempts to answer them.

It has been argued that, in order to move such an effort forward, a community of researchers, committed to a set of principles regarding evidentiary standards, must be developed and nurtured. We contend further that, in order to produce a base of scientific knowledge that is both rigorously derived and directly relevant to improving achievement in urban school districts, this community of inquiry must be expanded to include both scholars and practitioners in urban education.

Though a great deal of education research is produced every year, there is a genuine dearth of knowledge regarding how to address some of the fundamental challenges urban school districts face in educating children, working to close achievement gaps, and striving to meet the challenges of No Child Left Behind. Moreover, while there is a history of process-related research around issues affecting urban schools, relatively few studies carefully identify key program components, document implementation efforts, and carefully examine the effects of well-designed interventions in important programmatic areas on key student outcomes such as academic achievement. In sum, there is an absence of methodologically sound, policy-relevant research to help guide practice by identifying the conditions, resources, and necessary steps for effectively mounting initiatives to raise student achievement.

In order to address this need, the Council of the Great City Schools, through a grant from the Institute for Education Sciences, established the Senior Urban Education Research Fellowship (SUERF) program.

The Senior Urban Education Research Fellowship was designed to facilitate partnerships between scholars and practitioners focused on producing research that is both rigorous in nature and relevant to the specific challenges facing large urban school districts. We believe such partnerships have the potential to produce better, more practically useful research in at least three ways. First, by deepening researchers' understanding of the contexts within which they are working, the program may help them maximize the impact of their work in the places where it is needed the most. Second, by helping senior staff in urban districts become better consumers of research, we hope to increase the extent to which the available evidence is used to inform policy and practice, and the extent to which urban districts continue to invest in research. Third, by executing well designed studies aimed at the key challenges identified by the districts themselves, we hope to produce reliable evidence and practical guidance that can help improve student achievement.

[1] Council of the Great City Schools (2010). *Beating the Odds: An Analysis of Student Performance on State Assessment and NAEP.* Results from the 2008-2009 School Year. Washington, DC.

The primary goals for the Senior Urban Education Research Fellowship are to:

- promote high quality scientific inquiry into the questions and challenges facing urban school districts;
- facilitate and encourage collaboration, communication, and ongoing partnerships between senior researchers and leaders in urban school districts;
- demonstrate how collaboration between scholars and urban districts can generate reliable results and enrich both research and practice;
- produce a set of high quality studies that yield practical guidance for urban school districts;
- contribute to an ongoing discussion regarding research priorities in urban education; and
- promote the development of a "community of inquiry", including researchers and practitioners alike, committed to both a set of norms and principles regarding standards of evidence and a set of priorities for relevant, applied research in urban education.

The SUERF program benefitted greatly from the guidance and support of a Research Advisory Committee made up of experts and leaders from large urban school districts and the education research community. The committee included Dr. Katherine Blasik, Dr. Carol Johnson, Dr. Kent McGuire, Dr. Richard Murnane, Dr. Andrew Porter, and Dr. Melissa Roderick. This extraordinary group helped to identify and define the objectives and structure of the fellowship program, and we thank them for lending their considerable insight and expertise to this endeavor.

The following volume of the Senior Urban Education Research Fellowship Series documents the work of Dr. Catherine Snow and Dr. Joshua Lawrence, working in collaboration with the Boston Public Schools under the auspices of the Strategic Education Research Partnership (SERP). Both the research and reporting are the sole intellectual property of Drs. Snow and Lawrence, and reflect their personal experience and perspectives as education researchers.

Dr. Snow and SERP's work developing and implementing the Word Generation literacy intervention in Boston Public Schools illustrates the problem-solving potential of strong researcher-school district partnerships. The research team allowed district needs and priorities to drive the design of this innovative new literacy program which addresses the development of academic vocabulary -- a challenge identified by researchers and practitioners alike as a root cause of low student literacy and achievement levels.

The development team also took an important step in positioning the program as a cross-content area intervention. This recognition that teaching in any content area requires attention to literacy goes a long way toward building the type of collaborative work and culture necessary to transform schools into effective learning communities.

At the same time, Dr. Snow and her team document the challenges faced by education researchers in the process of designing, implementing, and evaluating a school-based intervention in a large urban school district. Of particular interest to school and district leaders, the report offers some insight into the characteristics and practices of schools that were able to implement Word Generation consistently and effectively. As we have seen in countless other studies and reports, the level and process of implementation largely determines the success of any given initiative, and we feel these "lessons learned" apply readily to other school-based reforms and programs being undertaken in school districts throughout the country.

The SERP team is currently pursuing more rigorous, systematic evaluation of Word Generation's impact on student learning, and we will continue to monitor the evolution and progress of the program. In the meantime, we hope you will find this "natural history of an intervention" interesting and relevant to your own work.

Michael Casserly
Executive Director
Council of the Great City Schools

ABOUT THE SENIOR URBAN EDUCATION RESEARCH FELLOW

Dr. Catherine Snow is the Patricia Albjerg Graham Professor at the Harvard Graduate School of Education. She received her Ph.D. in psychology from McGill and worked for several years in the linguistics department of the University of Amsterdam. Her research interests include children's language development as influenced by interaction with adults in home and preschool settings, literacy development as related to language skills and as influenced by home and school factors, and issues related to the acquisition of English oral and literacy skills by language minority children. Most recently she has focused on literacy development in adolescence, and interventions designed to improve adolescents' literacy skills. She has co-authored books on language development (e.g., *Pragmatic Development* with Anat Ninio) and on literacy development (e.g., *Is Literacy Enough?* with Michelle Porche, Stephanie Harris, and Patton Tabors), and published widely on these topics in refereed journals and edited volumes. Snow's contributions to the field include membership on several journal editorial boards, co-directorship for several years of the Child Language Data Exchange System, and serving as a member of the National Research Council Committee on Establishing a Research Agenda on Schooling for Language Minority Children. She chaired the National Research Council Committee on Preventing Reading Difficulties in Young Children, which produced a report that has been widely adopted as a basis for reform of reading instruction and professional development, and the National Research Council Committee on Developmental Assessments and Outcomes for Children. She is a past president of the International Association for the Study of Child Language and the American Educational Research Association. She heads the research activities of the Strategic Education Research Partnership's field site in the Boston Public Schools.

ABOUT THE RESEARCH PARTNERSHIP

In 2005 the Strategic Education Research Partnership (SERP) established its first field site, in the Boston Public Schools (BPS). The goal of a SERP field site is to improve the usability of educational research by functioning on a teaching hospital model, as a place where practice and research occur side-by-side, where practitioners and researchers together determine what the work should be, and where the complexities of student learning can be addressed within the context of attention to teacher learning and the organizational structure of schools and of districts.

SERP operates by devising and perfecting tools to ease the work of educators. One of the principles underlying SERP work (Donovan, Wigdor & Snow, 2003) is that those tools should be responsive to needs articulated by the practitioners themselves. This principle can be justified by the countless examples of evanescent educational reforms – new practices or materials that disappear quickly after introduction because they were imposed from outside (by researchers) or from the top (by district leaders) but were not seen by classroom teachers as responding to their needs. SERP proposed to start with a deep understanding of practitioners' needs and priorities, then to design tools that would be (and would be seen to be) responsive to those needs.

In addition to Boston, SERP currently operates field sites in San Francisco and in a subset of 4 smaller, inner-ring suburban districts from the Minority Student Achievement Network (MSAN). At the present time, participating MSAN districts are Ann Arbor (MI), Chapel Hill-Carrboro (NC), Evanston/Skokie 65 (IL), Madison(WI), and Shaker Heights(OH). The districts' leaders commit to regular meetings with key SERP researchers and staff to ensure the integration of the work with the district agendas and decision making.

Each field site has a different program of work. In Boston, the focus is on middle school literacy across the content areas. In San Francisco it is on middle school mathematics and science, and the literacy and language challenges of accessing content in those domains. In the MSAN site, the work is focused on algebra learning and on the engagement of students in academics at the transition to high school. While the foci differ, each site operates according to a common set of SERP principles:

1. The program of work is designed to address the problem(s) that the school district identifies as most urgent.
2. SERP recruits an interdisciplinary team of researchers, developers, and practitioners who are among the nation's most accomplished in the domain identified by the district.
3. Multiple lines of work are launched simultaneously to address the complexity of the challenges as they manifest in real school contexts.
4. Design work includes both researchers and practitioners at every stage. It attends from the start to designing for scale, and deliberately builds on prior work.
5. Interventions are subjected to rigorous scientific evaluation, providing solid evidence of their effect on student achievement.

In six short years, the field sites have been remarkably successful at deepening the engagement level and commitment of the school districts – even in times of transition – and at recruiting cooperative networks of researchers and practitioners who are among the best in the nation. Quality products, including assessments, instructional programs, pedagogical tools, and online professional development, have already begun to emerge from the work, attracting the interest of other districts facing similar challenges.

EXECUTIVE SUMMARY

PART I: DESIGNING THE WORD GENERATION PROGRAM

When the Strategic Education Research Partnership (SERP) began working with Boston Public Schools in 2005, the most pressing need articulated by the district was research and development in the area of middle school literacy. Thus SERP researchers undertook to specify more precisely what the middle school literacy problem in BPS was by interviewing middle school teachers and principals, by observing in classrooms, and by reviewing BPS test data. One universally noted challenge was vocabulary – students' ignorance of the meaning of the words they encountered in their texts. These challenging vocabulary items – words used across content areas, words characteristic of written language and academic texts, words students from non-English-speaking or low-literacy homes were unlikely to have heard from their parents – were not typically taught. This was a problem mentioned in particular by the science, social studies, and math teachers.

In response, Word Generation was designed to meet goals at three levels: 1) At the student level, the program would build knowledge of high frequency academic words, skills for spoken and written academic discourse, and knowledge about topics worthy of discussion; 2) At the teacher level, the program would assist in promoting regular use of effective strategies for teaching vocabulary, modeling comprehension, and promoting discussion usable in everyday instruction, and 3) At the school level, the program would help facilitate faculty collaboration across grades and across content areas.

However, in designing a literacy intervention tailored for the district, we also had to take into account various administrative contingencies and constraints faced by schools and the district. When all these various principles were integrated, we ended up with a program organized around weekly civic dilemmas selected to motivate students and to provide opportunities for authentic discussion. A full description of the program design is offered in Part I.

PART II: MEASURING IMPLEMENTATION

In 2007-2008, Word Generation was implemented in six Boston Public Schools. In order to gauge usability of the program and level of implementation, we used a range of methods tapping teacher and school sources at the six participating schools. From these various sources, we were able to identify three key features that impacted implementation of the Word Generation program at the school level:

Professional Development

Optimal professional development for adopting Word Generation involves prior planning and school-wide training. Prior to launching the intervention, we recommend a minimum of four hours (usually a morning and afternoon, perhaps staggered across a two-day period) of professional development. On-going professional development (two to three more school-based sessions) was also recommended.

Leadership and Accountability

Optimal implementation of Word Generation is both contingent upon and designed to enhance teacher accountability for student learning, high standards for student language and literacy skills, and openness to genuine discussion. These commitments, in turn, require strong leadership support and faculty collaboration.

Dedicated Staff

The appointment of a school-based Word Generation facilitator is also an important guarantor of high-level implementation. These facilitators oversee pre- and post-testing, monitor program implementation, provide school-based professional development, collect writing samples, and provide feedback to the program developers as to the challenges and levels of engagement by teachers and students.

As was to be expected, the six schools we worked with in 2007-2008 varied in the presence of these features, representing a wide range of "readiness" for new interventions. In Section II, we offer profiles of each of these school sites which emphasize differences between the kind of school that is poised to implement interventions and work collaboratively around issues of instruction and the kind that is not.

In 2008-2009, we adopted a new approach to thinking about implementation, using evidence from the student word-books to establish the intensity of implementation across content areas and the number of weeks of implementation across the school year.

These data reveal, first, that what is designed as a 24-week curriculum often becomes a 16- or 20-week curriculum, reflecting the often commented-on fact that little teaching occurs after the accountability assessments are administered in April. Second, the data suggests that there are differences across content areas in implementation. In general, the writing and focus word charts were most likely to have been completed, with math and science activities less widely implemented. This may reflect ongoing skepticism among math and science teachers about their responsibility for teaching vocabulary. Third, there are significant differences among the schools both in how many weeks they continued and in how thoroughly the cross-content-area model was followed. There is a strong correlation between effect sizes achieved in each school and the level of implementation found in student notebooks at those schools, and we expect these data to be a key component of future analysis.

PART III: EVALUATING PROGRAM EFFECTIVENESS

In addition to program design and implementation, the SERP team faced a key challenge in the area of program evaluation. In particular, we were interested in determining 1) whether the program helped students learn the target words, 2) whether gains in word knowledge were maintained over time and whether different subgroups of students showed similar patterns of gain and maintenance, and 3) if students who made gains in general purpose academic vocabulary did better on the state mandated ELA achievement test.

I. Measuring Vocabulary Development

First, to test whether the program helped students learn the target words, the team developed multiple-choice vocabulary tests with a selection of words from each week of the program, completed at the beginning and end of both the 2007-2008 and 2008-2009 school years. Section II provides a detailed discussion of the assessment challenges and limitations of the data yielded by this measurement tool. With these limitations in mind, the results demonstrate that students in Word Generation schools outperformed students in the comparison schools, although the effect sizes obtained from the second year are lower than those obtained the first year. We hypothesize that this diminished effect resulted from reduced fidelity and intensity of implementation in the second year.

II. Exploring Long-Term Impacts for Different Student Groups

While each set of pre-test and post-tests were designed primarily to assess knowledge of the words covered over the course of the corresponding year, 11 items taken from the first year's test were embedded in the second pre- and post-test. This allowed us to pinpoint the long-term effect of program participation on student vocabulary, and disaggregate this effect for students from English Only homes (EO), students from Language Minority homes (LM), and Limited English Proficient (LEP) students.

The results of this analysis suggest that students from language-minority homes who participated in the program made strong gains – gains that put their scores above those of EO students in comparison schools – from the intervention. Furthermore, they maintained those gains relative to comparison students even a year later. Students from English-speaking homes also made gains relative to the comparison group and maintained them

across the course of the study. However, LEP students did not show comparative benefits from participation in the Word Generation program; their rate of growth continued to parallel that of their LEP peers in the higher achieving comparison schools, with no narrowing of the gap.

III. Examining the Relationship of Word Generation Participation to MCAS Scores

Finally, we conducted an exploratory analysis to determine whether participation in Word Generation had any relationship to performance on the MCAS. Using regression analysis, we constructed a model with MCAS scores in April, 2008 as the outcome, using gender, treatment status, pre-test and post-test scores as predictors. Results indicate that improvement from Word Generation pre- to post-test did indeed predict MCAS scores for Word Generation students, but not for students in comparison schools. We think it highly plausible (though subject to further confirmation) that the discussion, deep reading, and regular writing activities incorporated into Word Generation helped students perform better, particularly on those MCAS items requiring reading comprehension and open responses.

DISCUSSION AND CONCLUSION

The findings of this quasi-experimental study were highly informative, both about the potential of innovative approaches to support students' academic progress and about the challenges to an optimal implementation and evaluation of a literacy program. The report concludes with a discussion of ongoing work in the development and evaluation of Word Generation and reflections on working collaboratively within urban districts.

INTRODUCTION

INTRODUCTION

THE DEVELOPMENT OF WORD GENERATION

When SERP began working with Boston Public Schools in 2005, the most pressing need articulated by the district was research and development in the area of middle school literacy. For Thomas Payzant, then the superintendent, and indeed for much of the research and practice community, the failure of students to make ongoing progress in reading after the primary grades was puzzling, and it was alarming furthermore that so many of them ended up disastrously unprepared for the challenges of content area learning in high school. In the years since, the phenomenon identified by Superintendent Payzant has received increasing attention as the 'adolescent literacy crisis' (see, for example, www.carnegie.org/literacy), but in 2005 the exact nature of the challenge remained obscure.

Thus SERP researchers undertook to specify more precisely what the middle school literacy problem in BPS was by interviewing middle school teachers and principals, by observing in classrooms, and by reviewing BPS test data. Our goal was to identify the need so we could design tools to help teachers and schools address that need, and to understand how teachers themselves defined the most urgent problem.

Not surprisingly, teachers offered many reasons for the poor literacy skills of their students. Poor inferencing skills, low stamina, lack of motivation, distractions of television and videogames, lack of parental support, peer pressure, and other factors were all mentioned. But one universally noted challenge was vocabulary – students' ignorance of the meaning of the words they encountered in their texts. This was a problem mentioned in particular by the science, social studies, and math teachers. Because they didn't know what many key words meant, teachers reported, the students could read paragraphs from their texts correctly and fluently, but at the end they couldn't tell you what they said.

It seemed, then, that an effort to support students' vocabulary development might contribute to their literacy success, and might also be recognized by BPS teachers as a response to the needs they identified. BPS teachers did almost universally teach vocabulary, of course, but their instruction was focused on the vocabulary items relevant to their own content areas. Additional challenging vocabulary items – words used across content areas, words characteristic of written language and academic texts, words students from non-English-speaking or low-literacy homes were unlikely to have heard from their parents – were not typically taught.

Thus, the SERP research team decided to start there, with a vocabulary program focused on all-purpose words useful across all the content areas. In addition, recognizing that these words are as likely to occur in science as in math as in social studies, we decided to incorporate activities for these content area teachers to implement, and not leave vocabulary as the sole province of the English Language Arts (ELA) teacher.

These general principles drove the design for the Word Generation. In this report, we summarize what we have learned from the work done on Word Generation in collaboration with the Boston Public Schools, share the challenges we faced in the design, implementation, and evaluation of the program, and present some guidance based on this experience to others interested in partnership-based educational research and development.

PART I:
DESIGNING A LITERACY INTERVENTION FOR BOSTON PUBLIC SCHOOLS

PART I: DESIGNING A LITERACY INTERVENTION FOR BOSTON PUBLIC SCHOOLS

As a cross-content vocabulary program designed to develop all-purpose, high leverage vocabulary and academic language, Word Generation addresses the needs of middle school students struggling with comprehension of their texts. These ideas provided the basic design principles for Word Generation: all-purpose academic words, and cross-content area instruction. 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

However, in designing a literacy intervention tailored for the district, we also had to take into account the specific achievement levels and needs of BPS students, schools, and the district. While a majority of 6th-8th graders in many Boston middle schools fall into the category of struggling reader, they study in classrooms with average and good readers.

Thus, though the overarching goal of Word Generation is to employ systematic vocabulary instruction to improve student achievement in schools serving large concentrations of low-income children and English language learners, we also had to make the program engaging and productive for more successful readers.

Yet another set of design principles derived from BPS administrative contingencies and constraints: not more than 15 minutes a week to be devoted to the program in math, science, or social studies, limited time for professional development with teachers, the requirement of some common planning time at the school level, and the relevance of math activities to math standards and the state accountability assessment (MCAS) formats.

Furthermore, we discovered in our pilot work that middle schools chose to implement the program school-wide. In other words, the same curriculum was used with 6th-8th graders. Thus we had to select topics and tasks that were appropriate across that range, and that could be made relevant to all the content areas. To design and implement an effective language intervention that crosses grade levels and content areas is a challenging enterprise; doing this for use in underperforming schools with low levels of academic achievement and sometimes incoherent organizational structures is even harder. Interventions work best if they initially receive wide support by leadership and practitioners and they clearly address a district or school-identified concern. They work even better in schools where there are shared commitments and responsibilities for teaching and learning.

The middle school literacy challenge in BPS was particularly acute for some schools and some students. In particular, the district noted that both English language learners (ELLs) and native English-speaking students from low-income families were faring poorly on district and state assessments because of their limited vocabularies; classroom practitioners confirmed that because students lacked academic language and vocabulary they did not know many of the words presupposed in content-specific texts. This limited their ability to comprehend or learn from these materials.

In response, Word Generation was designed to meet goals at three levels: 1) At the student level, the program would build knowledge of high frequency academic words, skills for spoken and written academic discourse, and knowledge about topics worthy of discussion; 2) At the teacher level, the program would assist in promoting regular use of effective strategies for teaching vocabulary, modeling comprehension, and promoting discussion usable in everyday instruction, and 3) At the school level, the program would help facilitate faculty collaboration across grades and across content areas. Effective implementation of Word Generation is highly dependent on this third dimension-- specifically, the capacity of personnel within each school to work collaboratively and with accountability around issues of instruction. Moreover, distributing responsibility for implementing the program across all content-area teachers was designed to reduce the burden on any single teacher or content area while providing recurrent exposures to the target words, in a variety of semantic contexts, and to ensure that all the various content area teachers could learn and practice research-based strategies for teaching vocabulary and academic language.

When all these various principles were integrated, we ended up with a program organized around weekly civic dilemmas selected to motivate students and to provide opportunities for authentic discussion. Each week was organized around a topic selected to be engaging for young adolescents and to generate a genuine question – an issue on which a number of points of view can be plausibly defended. Sample topics included some very close to students' lives (e.g., Should school uniforms be required? Should rap music be censored? Should schools stop selling junk food? Should passing a standardized test be a high school graduation requirement?) and others that were more remote, but related to topics of national interest (Is animal testing of drugs and cosmetics necessary? Should secret wiretapping be legal? Should advertising of prescription drugs on television be allowed? Is the death penalty fair?). Words taught explicitly in the program include ones needed for making and evaluating arguments (e.g., evidence, support, claim, affirm, deny), for structuring discourse (e.g., thus, moreover, nonetheless), for referring to abstract entities (e.g., factor, process, phenomenon, theory), for hedging claims (e.g., evidently, disproportionately), and so on.

Each dilemma was introduced on Monday by the ELA teacher, using a passage of about 300 words in which five target academic words were embedded. On Tuesday, Wednesday, and Thursday, following a schedule determined by each school, topic-related activities in math, science, and social studies were implemented. Math and science problems related to the week's dilemmas ensure that students have opportunities to hear the words in a variety of settings, where furthermore discipline-specific meanings (e.g., factor in math, process in biology) can be explained. The week's dilemma is further explored in a debate or issue-focused discussion staged in social studies class. On Friday, the ELA teacher assigned a brief 'taking a stand' essay, in which students were asked to select and defend their position on the dilemma of the week.

Complete information about the topics, words taught, and tasks is available at www.serpinstitute.org/wordgeneration, where it is also possible to view clips of teachers implementing the activities as well as interviews with students, teachers, and principals.

PART II:
MEASURING IMPLEMENTATION

PART II.
MEASURING IMPLEMENTATION

PROGRAM IMPLEMENTATION 2007 – 2008

In 2007-2008, Word Generation was implemented in six Boston Public Schools (all schools are identified with pseudonyms). Four of these were middle schools and two served children from Kindergarten through the eighth grade, but used the Word Generation program only in their 6th, 7th, and 8th grade classes. Reilly and Westfield Middle Schools had piloted 12 weeks of Word Generation in 2006-2007, and thus were in their second year of implementation, while Mystic, Occidental (both K-8) and Gorham and Mercer Middle Schools adopted the program for the first time in September 2007. All the middle schools adopted the program using a whole-school model (all grades using the same curricular units) except Gorham, where it was used only in the substantially separate special education unit.

In order to gauge usability of the program and level of implementation, we used a range of methods tapping teacher and school sources at the six participating schools. At the teacher level, we used a number of sources of information: school-based participation in professional development opportunities (prior to launch in the school and throughout the school year); participation in cluster and grade-level team meetings devoted to understanding and improving the intervention; classroom observations; informal and structured interviews with teachers; teacher feedback surveys; and video-taping of exemplary teaching. At the school level, we also used a number of sources of information to gauge institutional commitment: the support from and involvement by each principal in disseminating and overseeing the intervention; practical provisions (e.g., scheduling time for professional development, scheduling meeting and planning time for teachers throughout the school year; making time and space available for school staff to organize assessment and implementation schedules); and monitoring in an informal way how universally the program was being accepted and used.

Feedback from Teachers on Word Generation Successes and Challenges

Of approximately 200 teachers participating in Word Generation during 2007-2008, 62 teachers returned completed on-line surveys for the first half of the intervention (units 1 through 12), and 81 teachers returned questionnaires for the latter half of the program (units 13 through 24). These questionnaires were designed to garner teacher feedback on the successes and weaknesses of the program, their students' levels of engagement and their own thoughts about how to improve the intervention. Many wrote that their students were able to think critically about the controversial topics embedded in the curriculum as well as write about them and debate the issues. One teacher reported that her students "engaged with many of the issues, resulting in interesting discussions and better developed essays." Focusing on program implementation, several teacher participants wrote of the cohesion that emerged across content areas and for the school as well.

There were also comments that attested to the challenges of implementing the program in some schools. For example, teachers complained that implementing the Word Generation curriculum often took more than the 15-20 min allotted to it, especially if the students became highly engaged in the topic. Teachers noted that the Word Generation topics were not aligned with their curricular topics—sometimes Word Generation introduced topics not covered elsewhere, and sometimes it introduced topics relevant to the curriculum but at disparate times. Some math, science, and social studies teachers rejected the notion that they should be responsible for teaching all-purpose vocabulary, suggesting that this should be the ELA teacher's task. Finally, there were occasional complaints about the topics themselves, including objections to our treatment of climate change ("a myth"), reluctance to broach issues such as sex education or sexting, and worry that some topics (causes of diabetes, stem cell research) went beyond their own knowledge base.

Features Impacting Implementation at the School Level

Professional Development

Optimal professional development for adopting Word Generation involves prior planning and school-wide training. Prior to launching the intervention, we recommend a minimum of four hours (usually a morning and afternoon, perhaps staggered across a two-day period) of professional development. These hours are devoted to presenting background research on vocabulary teaching and learning, introducing the Word Generation approach, viewing video-clips of exemplary implementation of the program by other BPS practitioners, and providing opportunities for hands-on practice with program materials and activities. Adopting schools were strongly encouraged to send teams to a Word Generation institute offered in the summer of 2007. The institute offered opportunities for school-based staff and leadership a) to become familiar with the program design, materials, and activities, b) to provide the developers with feedback and recommendations for improving the content of the intervention, and c) to organize school-level assessment, professional development, and implementation schedules. On-going professional development (two to three more school-based sessions) was also recommended and most schools established several dates throughout the academic year for feedback and professional development sessions. In Table 1 we summarize the degree of participation in professional development opportunities by each of the six schools using Word Generation in 2007-2008.

Leadership and Accountability

Optimal implementation of Word Generation is both contingent upon and designed to enhance a set of shared understandings and commitments at the school level. These commitments include teacher accountability for student learning, high standards for student language and literacy skills, and openness to genuine discussion. These commitments, in turn, require strong leadership support, faculty collaboration, and opportunities for regular cluster, grade or content level meetings where implementation schedules can be reviewed, materials can be previewed, and team building activities can occur.

Dedicated Staff

The appointment of a school-based Word Generation facilitator is also an important guarantor of high-level implementation: these facilitators oversee pre- and post-testing, monitor program implementation, provide school-based professional development, collect writing samples, and provide feedback to the program developers as to the challenges and levels of engagement by teachers and students.

Implementation by School

As was to be expected, the six schools we worked with in 2007-2008 varied in the presence of these features (see Table 2). In the next section we provide a quick portrait of each of the six schools.

All of the participating Word Generation schools had a high percentage of students living in poverty (ranging from a low of 79% free and reduced lunch-eligible students to a high of 91%), substantial levels of students with special education (SPED) designations (between 16% and 33%), and many students from second language homes (from 32% to 70%) (see Table 3). All schools (except one) offered Sheltered English Immersion classrooms to their limited English proficient (LEP) students, and all these sheltered classrooms implemented Word Generation, albeit with a range of modifications.

PART II.
MEASURING IMPLEMENTATION

TABLE 1: PARTICIPATION IN PROFESSIONAL DEVELOPMENT BY SCHOOL, 2007-2008

SCHOOLS	OVERALL LEVEL OF PARTICIPATION IN PD	PARTICIPATION IN *WORD GENERATION* SUMMER INSTITUTE	TOTAL HOURS OF PROFESSIONAL DEVELOPMENT	ON-GOING PROFESSIONAL DEVELOPMENT?
Reilly	High	Yes	18	Yes
Westfeild	Medium	Yes	16	Yes
Mystic	Medium	No	4	No
Mercer	Low	No	2	No
Occidental	High	Yes	12	Yes
Gorham	N/A	No	2	No

TABLE 2: ORGANIZATIONAL FEATURES OF PARTICIPATING SCHOOLS, 2007-2008

SCHOOLS	YEARS OF IMPLEMENTATION	BUILDING CAPACITY	LEADERSHIP SUPPORT	ORGANIZATIONAL COHERENCE
Reillly	2	High	High	High
Westfeild	2	Low	Low	Low
Mystic	1	High	High	High
Mercer	1	Low	High	Low
Occidental	1	High	High	Low
Gorham*	1	Low	Low	Low

* Learning and Adaptive Behavior (LAB) cluster, substantially separate program for students with special education designations

Reilly Middle School

Reilly Middle School is one of those schools poised to implement any intervention successfully because of the personnel's capacity to work collaboratively and because of collective responsibility for student learning. Reilly easily had the highest level of cohesion and internal accountability of all participating Word Generation schools. Leadership was strong, there was cooperation across the content areas under the solid direction of the literacy coach/Word Generation facilitator, and there was a climate of trust and shared accountability across content areas and grade levels.

Reilly saw itself as a learning organization with an established capacity to improve instruction. This perception was supported by their experience successfully implementing district-initiated curricular packages and their relatively high Massachusetts Comprehensive Assessment System (MCAS) scores for a school with their demographic profile (see Table 4). Their schedule allowed for regular meetings of grade- and content-area teams, and for targeted, ongoing professional development led by the full-time literacy coach.

The school's commitment to the program originated with the instructional leadership team, especially the principal and the literacy coach. It was sustained by teacher satisfaction with the program activities and outcomes as well as student enthusiasm for the topics and opportunities for discussion. In addition, in order to teachers motivated a system for providing them with in-service credits was devised by the literacy coach. Receiving the credits was contingent on full implementation of the program and completion of three questionnaires responding to the activities and the topics.

Westfield Middle School

Westfield Middle School is a chronically underperforming school that was under threat of closure in 2007-2008. Westfield faced many challenges common to high poverty schools with poor academic track records: limited organizational capacity, high teacher turnover, low teacher morale, and weak leadership. Westfield had limited success securing the commitment of all content-area teachers across the grade levels to implement the vocabulary intervention. However, individual teachers were committed to daily implementation of the program, and the seventh grade team operated in a more cohesive way than other grade level teams.

In an effort to support Westfield's use of Word Generation, researcher teams met with grade level teachers in cluster meetings, shared promising data with staff, and attempted to increase the level of participation and trust with the principal. The principal had given her tacit support to the program but often failed to follow through on commitments. Although we met and communicated on several occasions to establish the assessment calendars, these were not adhered to. For example, the entire eighth grade left the building for a field trip on the date of the scheduled post-test, complicating the process of obtaining end-of-year data.

However, the principal gradually became more active in her support for the program; she began to observe Word Generation lessons in action and indicated she wanted to increase fidelity to the program practices at Westfield.

To provide further support, an ELA teacher who was very committed to the program was hired as the school-based facilitator to distribute materials and demonstrate Word Generation lessons for her colleagues. In addition, a graduate student working with the Word Generation curriculum was assigned to the school. During multiple visits and interviews with teachers, the student learned that scheduling of Word Generation lessons within clusters had proved a major stumbling block. Math, science, and social studies teachers felt unable to do their lessons when English language arts teachers had not implemented the launch activity to introduce the week's topic and words. Moreover, a number of teachers considered the program optional. Added to these challenges, the principal was distressed upon receiving survey results showing that the school and her staff had low internal coherence and low regard for her leadership.

Nonetheless, the principal did invite us to conduct a school-wide professional development workshop, which she attended. This workshop emphasized not only the importance of academic vocabulary and academic discussion, but also the importance of creating workable schedules. Pointedly, the workshop also included a hands-on activity directing teachers to create schedules for each cohort of students. When some teachers balked at the activity, the principal stressed the program's importance and urged them to find solutions to their scheduling problems. Teachers then transferred the schedules to charts which were shared with the group. These schedules, collected by the Word Generation team, were organized and emailed back to the principal.

Subsequently, school support for Word Generation continued to improve. In January, when student identification numbers were requested to facilitate monitoring of student progress, the principal personally photocopied the list. School hallways boasted two Word Generation bulletin boards, and while neither bulletin board contained student work, their existence increased Word Generation's visibility, and was indicative of its increased importance onsite.

Along with indications of the program's increased importance, however, were areas of concern. It was troubling, for instance, that not all grade cohorts were tested at the same time, and that social studies teachers, unsure how to conduct a debate, tended to neglect this component. The SERP team met and discussed these issues and planned next steps for scheduling; there was also a plan for providing teachers with a simpler approach to the debate format, a discussion in which students were guided to agree, disagree, and extend each other's comments. Because the principal identified the debate activity with her goal to increase academic discussions through "accountable talk," she agreed to

PART II.
MEASURING IMPLEMENTATION

TABLE 3: DEMOGRAPHIC INFORMATION ABOUT PARTICIPATING SCHOOLS, 2007 - 2008[2]

SCHOOL	PERCENTAGE OF STUDENTS FROM LM HOMES	PERCENTAGE OF STUDENTS WITH IEPS	PERCENTAGE OF STUDENTS RECEIVING FREE OR REDUCED PRICE LUNCH
TREATMENT SCHOOLS			
REILLY	31.9	26.3	82.7
MERCER	43.9	25.4	87.4
WESTFIELD	48.8	27.1	78.8
MYSTIC	69.7	16.1	89.7
OCCIDENTAL	51.7	32.6	90.7
COMPARISON SCHOOLS			
GARFIELD	54.0	31.7	81.0
JEFFERSON	35.8	30.7	78.1
UXTON	61.4	19.7	89.8
AVERAGE SCORES			
TREATMENT	49.2	25.5	85.8
COMPARISON	50.4	27.4	83.0

TABLE 4: ELA MCAS SCORES OF PARTICIPATING SCHOOLS, 2007 - 2008[3]

SCHOOL	ELA MCAS 2007				ELA MCAS 2008				
	W*	NI	P	A	W	NI	P	A	SCORE
TREATMENT SCHOOLS									
REILLY	10.7	33.7	52.3	4.0	6.3	33.3	56.0	4.7	242.4 (12.4)
MERCER	16.3	40.3	43.0	1.0	18.7	35.3	42.3	4.0	238.2 (14.2)
WESTFIELD	27.0	42.0	30.0	0.7	27.7	45.3	27.3	0.0	233.7 (11.5)
MYSTIC	2.3	30.3	63.3	4.0	5.7	30.0	58.7	5.7	241.1 (13.0)
OCCIDENTAL	35.3	38.7	26.0	0.0	42.0	33.3	24.3	0.3	234.0 (13.5)
AVERAGE	18.3	37.0	42.9	1.9	20.1	35.5	41.7	2.9	239.3 (13.5)
COMPARISON SCHOOLS									
GARFIELD	3.0	26.3	68.3	2.0	5.3	30.0	61.3	3.7	240.3 (12.1)
JEFFERSON	4.0	39.0	51.0	7.0	9.0	45.5	37.0	8.0	242.2 (16.2)
UXTON	21.3	39.0	37.0	3.0	26.3	32.7	39.0	2.3	239.4 (10.1)
AVERAGE	9.4	34.8	52.1	4.0	13.6	36.1	45.8	4.7	240.3 (12.3)

* W = warning, NI = needs improvement, P = proficient, A = advanced. Only P and A are considered passing.

encourage it. She also praised several aspects of the program. Specifically, she found the new bound book format teacher-friendly, the program's focus on current issues valuable, and the similarity of the Word Generation essay to the MCAS free response helpful. In fact, she announced plans for cluster meetings so teachers could examine student essays.

Increased assistance to Westfield likely bolstered principal support, fidelity of implementation, visibility of the program, and effective partnership. During a cluster meeting, the principal commented, "we improved our implementation this year, we intend to improve it even more next year."

[2] Adapted from Snow, C., Lawrence, J., & White, C. (2009).

[3] Adapted from Snow, C., Lawrence, J., & White, C. (2009).

Mystic K-8

The Mystic is a small neighborhood school that serves a largely Latino and language minority population of students. The Mystic has been lauded within Boston for its effectiveness in serving its Latino and ELL population; the almost exclusively Anglo teaching staff nurtured a sense of community with their students, for example, by studying Spanish and involving students in helping them learn it. The principal, who was highly effective, had identified vocabulary as a particular challenge for her students, and had previously introduced an intensive vocabulary program in grades 1-4. After her staff declared Word Generation was too challenging to implement and declined to do it, she championed the program and insisted it be taught.

The vice principal, meanwhile, took on the role of Word Generation facilitator and oversaw the implementation and testing. Mystic had only two content area teachers per grade in the middle grades – one for ELA and social studies, and one for science and math. Thus, the scheduling of the Word Generation activities involved only a few individuals and did not pose a problem. Teacher commitment to the program may have been enhanced as well by their recognition that their students, almost all of whom came from Spanish-speaking homes, struggled in particular with vocabulary. Mystic was the only school of the six that completed the entire 24 week curriculum. Overall, quality of implementation was very high, in particular in classes taught by a stellar science teacher.

Mercer Middle School

The organizational capacity for undertaking certain tasks is evident at Mercer Middle School, a large middle school (almost 700 students) headed by an action-oriented interim principal. This principal enthusiastically volunteered her school as a Word Generation site but struggled subsequently to convince her staff to share her enthusiasm and commitment to Word Generation. For the required initial professional development session, she was able to recruit classroom teachers only to a two-hour session, which most attended grudgingly. Even within the context of staff resistance to the adoption of the program, she was able to organize the assessment calendars and actual testing of impressive numbers of students, including Mercer's large English language learner population. She managed the distribution and dissemination of the program materials but was not able to oversee implementation at the classroom level in any systematic fashion.

One teacher lamented the lack of cohesion among the Mercer staff by writing this comment in the teacher feedback survey:

[There is] "no consistency; not everyone participating in WG the same way. [The] pre-test was not given with consistency (ie – kids were allowed to work together on pre-test!); teachers and admin were unclear on the goals of Word Gen. Truly, the adults here, it was quite clear, did not know/understand the purpose of the program – [that it was] not a laundry list of vocab words but rather, a way to frame academic language and high frequency words and access them through various lenses. If adults didn't get this, it was not taught as such."

The principal's interim status likely limited her authority to make the kinds of decisions necessary for securing a more effective roll-out and subsequent implementation. Beyond a superficial organizational capacity, the school's capacity to improve instruction through this program was quite low, although some classroom teachers were able to work together and carve out collaborative pockets of inspired instruction. Although Word Generation is designed to foster the conditions necessary for effective implementation, we find schools such as Mercer with very low capacity need more organizational supports than the program by itself can provide.

Occidental K-8

Occidental serves large numbers of low-income children (90.7%), many of whom come from language minority homes (51.7%). This school has failed to meet state standards, has been designated as underperforming, and has been in corrective action for several years. Only about a quarter of the student body performed at a proficient level on the 2007 ELA MCAS (see Table 4). Occidental has also experienced difficulty with leadership and leadership retention; in the fall of 2007 a

PART II. MEASURING IMPLEMENTATION

new principal replaced the out-going administrator after several contentious years, only to leave himself at the end of the academic year for a more promising position.

However, this school, even within a difficult and sometimes depressing school climate, was extremely effective in its implementation of the program. The successful implementation can be attributed to the involvement of leadership, an exceptionally committed literacy coach (and Word Generation facilitator), and strong teacher content-area teams. The principal and teacher teams participated in initial Word Generation professional development sessions, and then continued planning and working together on a weekly basis to establish implementation schedules and to share their successes and challenges with the program. On various occasions, these teams requested support from members of the Word Generation professional development team on how to better implement aspects of the program they found especially challenging. These supplementary professional development sessions were mutually productive, as they offered the opportunity for the Word Generation team to receive feedback on all aspects of the program, including many recommendations to improve the content of the program.

Most teachers were very positive about the program and described Word Generation's impact on word learning, writing quality, and engagement by their students. The Word Generation facilitator provided teachers with very strong direction as well, overseeing implementation, collecting writing samples, and providing feedback to the program developers about the needs of teachers and students. Because of its cohesive organizational structures within the middle-grades program, Occidental offered an excellent example of systematic vocabulary instruction put into action.

Gorham Middle School

The Gorham Middle School has not made AYP since 2003 and in 2007-2008, was in its second year of restructuring under NCLB. The school is characterized by high teacher turnover, a high percentage of special education students (34%), and benign but largely ineffective leadership. Although many programs and initiatives have been adopted at the school, the emphasis for the past few years has focused less on instruction than on establishing discipline and improving school climate, both of which have indeed improved.

The principal is well-intentioned but has yet to build the kind of school-level trust and commitment to student learning necessary for programmatic success.

A new energetic director of Special Education requested to use Word Generation in a segregated special education setting. Five volunteer teachers generated their own goals for the program, stating that it would be used to build "expressive and receptive vocabulary in speech and text" in their mixed grade Learning and Adaptive Behavior (LAB) Cluster. We were given two hours for the introductory professional development session and a graduate student from the Harvard Graduate School of Education (HGSE) and former special education teacher was assigned to provide assistance to this small group of teachers and students. She conducted structured observations, collected and analyzed writing samples, and interviewed students about their opinions of the program.

Interestingly, these interviews provided a useful look at actual implementation; students reported that implementation was not optimal. One student suggested that "teachers stop giving students the answers" and that "they should use the program more, so students have more time to practice." Some of the content from student interviews also gave evidence that students have strong, emotional ties to the intervention. When interviewed, it was also found that the participating teachers were not aware of their students' actual reading levels. There were modest gains overall in students' knowledge of the target words and greater gains in classrooms where implementation was more faithful.

PART II.
MEASURING IMPLEMENTATION

Summary

The six brief school portraits presented here emphasize differences between the kind of school that is poised to implement any intervention and work collaboratively around issues of instruction and the kind that is not. Schools in the first category have high levels of internal accountability (Fuhrman & Elmore, 2004). Leadership and staff collectively decide on high-priority commitments, and then hold each other accountable for follow-through on those commitments. In other words, these schools function as learning organizations.

The six Boston schools that implemented Word Generation in 2007-2008 represented a wide range of "readiness" for new interventions, as indicated by measures of their internal accountability and capacity for the kind of collaborative work necessary for effective implementation of the program. All six Word Generation schools were volunteer adopters with similar demographics and challenges; however, their capacity for implementing the program optimally differed greatly in areas of leadership, organizational coherence, commitment to professional development, and teacher buy-in.

PROGRAM IMPLEMENTATION 2008 – 2009

In 2008-2009, we adopted a new approach to thinking about implementation, one that was more objective than the brief case studies done in 2007-2008 but also much more time-consuming and labor intensive. In effect, we used evidence from the student word-books to establish which elements of the program were actually taught during which week. Ideally, students show their work in their word-books by filling in the focus word chart (ELA), working the math problem, filling in the cloze passages used for science, possibly making notes on the social studies page, and completing the taking-a-stand essay. We collected student wordbooks and coded them as an indication of the intensity of implementation across content areas, and the number of weeks of implementation across the school year. Figures 1 and 2 present example implementation data from two schools.

Each data point in these figures reflects how many students in that school showed evidence in their Word Generation Wordbooks that they had done the activity in that content area during that week. Thus, for example, in Week 2 in at Reilly Middle School (Figure 1), about 275 students did the writing activity, but only about 140 showed evidence of having done the science activity. Week 7, on the other hand, showed a decline in implementation across all content areas, possibly reflecting some external force such as a snow-shortened week or a school-wide assessment activity.

These data reveal, first, that what is designed as a 24-week curriculum may be transformed within the schools into a 16- or 20-week curriculum, reflecting the often commented-on fact that little teaching occurs after the accountability assessments are administered in April.

Second, the data suggests that there are differences across content areas in implementation. In general, the writing and focus word charts were most likely to have been completed, with math and science activities less widely implemented. This may reflect ongoing skepticism among math and science teachers about their responsibility for teaching vocabulary.

Third, there are significant differences among the schools both in how many weeks they continued and in how thoroughly the cross-content-area model was followed. There is a strong correlation between effect sizes achieved in each school and the level of implementation found in student notebooks at those schools, and we expect these data to be a key component of our year three analysis. We also expect that they will inform our work with the Word Generation program in other districts.

We are still struggling with more efficient ways to code and aggregate these data, but we have noted that reporting them to school leaders serves as useful input to their understanding of how teachers are responding to their plans for Word Generation use.

FIGURE 1: IMPLEMENTATION OF WORD GENERATION CONTENT AREA ACTIVITIES BY WEEK AT REILLY MIDDLE SCHOOL DURING 2008-2009, BASED ON STUDENT NOTEBOOK EVIDENCE (N = 364).

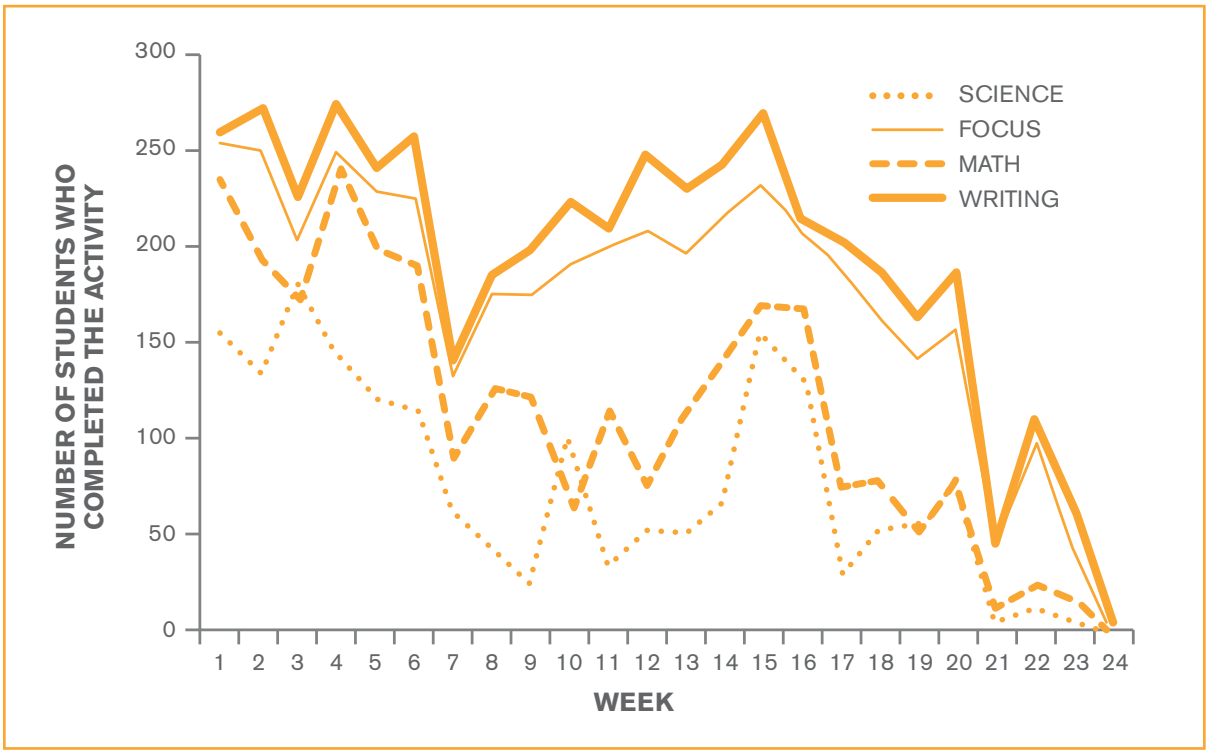
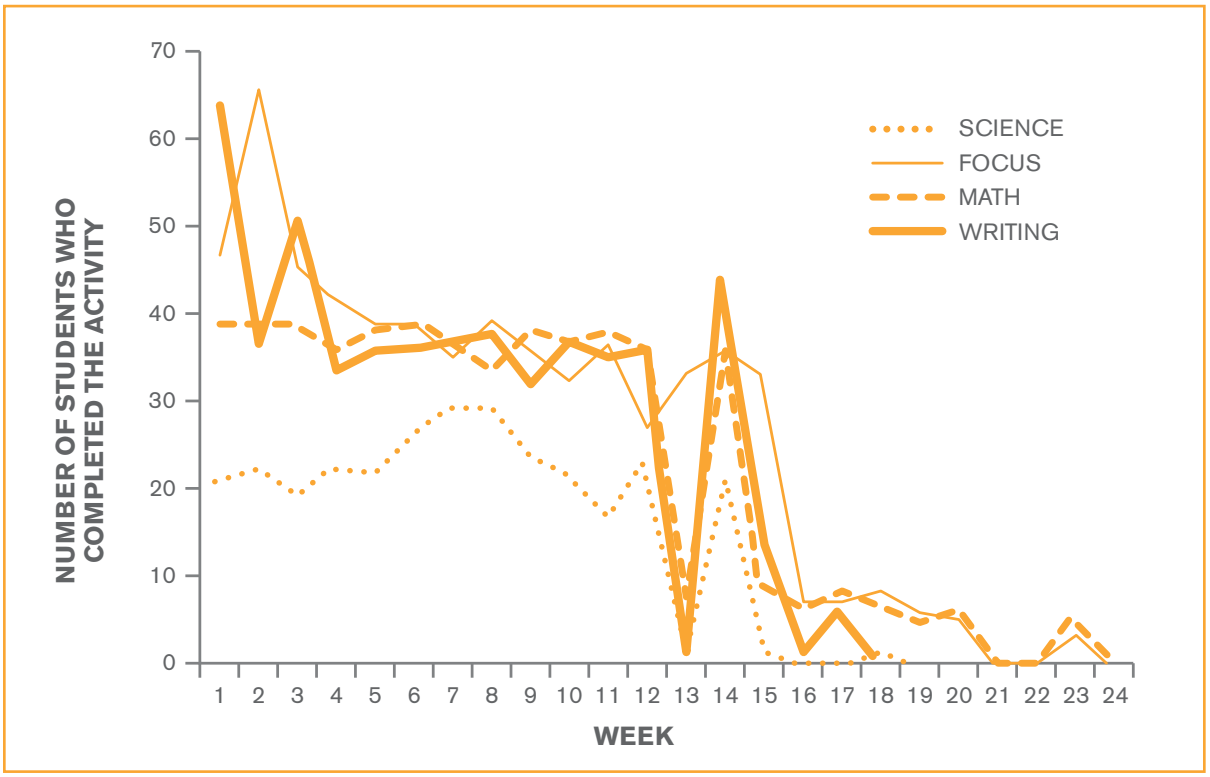


FIGURE 2: IMPLEMENTATION OF WORD GENERATION CONTENT AREA ACTIVITIES BY WEEK AT MYSTIC K-8 DURING 2008-2009, BASED ON STUDENT NOTEBOOK EVIDENCE (N = 70).



PART III:
EVALUATING PROGRAM EFFECTIVENESS

PART III.
EVALUATING PROGRAM EFFECTIVENESS

In addition to program design and implementation, the SERP team faced a key challenge in the area of program evaluation. In particular, we were interested in determining 1) whether the program helped students learn the target words, 2) whether gains in word knowledge were maintained over time and whether different subgroups of students showed similar patterns of gain and maintenance, and 3) if students who made gains in general purpose academic vocabulary did better on the state mandated ELA achievement test.

MEASURING VOCABULARY DEVELOPMENT

First, to test whether the program helped students learn the target words, the team developed a program-specific vocabulary test to be administered to students from the six schools that implemented the Word Generation program and three schools recruited by BPS as comparison cases (one school, the Gorham, implemented the program only in special education classrooms, so its results are not included in the general analysis presented here).

In the first year, this test included 48 multiple choice questions that randomly sampled two of the five words taught each week. Both pre- and post-test data were collected for 697 students in five treatment schools and 319 students in three comparison schools. All students in the treatment schools received the intervention; those contributing to the analysis reported here were the subsample that had completed usable test forms at both pre- and post-test.

There were 349 girls and 348 boys who met these criteria in the treatment schools, and 162 girls and 157 boys in comparison schools. Of these, 438 were classified as Language Minority (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 3 in the previous section, the vast majority of students in both treatment and comparison schools were from low-income homes. Furthermore, the data reported in Table 4 suggest that the comparison schools were performing better than the treatment schools at the start of the study, and that impression was confirmed by disparities in perfomance on the curriculum-specific pre-test.

Assessment Challenges and Data Limitations

Of course, because the implementing schools were those that volunteered for the program, selection effects must be taken into account in interpreting the findings.

In addition, we encountered two major challenges in the administration of the tests in the first year of the quasi-experimental study that have implications for the validity of the data.

- *Pacing difficulties.* 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 four items from our analysis of both pre- and post-test.
- *A time lapse in the administration of the pre-test in treatment and comparison schools.* The pre-test was successfully administered to students in all the treatment schools in October 2007, before the introduction of Word Generation materials. Yet because of difficulty recruiting the comparison schools, their pre-tests were not administered until January. The post-test (identical to the pre-test 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 post-testing in the two groups of schools, we present data on words learned per month as well as total words learned.

Findings, 2007-2008

With these data limitations in mind, the results were promising in the first year of study.[4] Descriptive results suggest that students in the Word Generation program learned approximately the number of words that differentiated 8th from 6th graders on the pre-test—in other words, participation in 20-22 weeks of the curriculum was equivalent to two years of incidental learning.

Unfortunately, the relative improvements in the Word Generation schools will be exaggerated by the differences in timing of the pre-test. Table 5 presents both the total pre- to post-test improvement, and also the improvement divided by the number of months between pre- and post- tests (8 months for treatment schools, 5 months for comparison schools). The results shown in Table 5 demonstrate that Word Generation schools outperformed the comparison schools even when the amount of time between tests is taken into consideration. The last column of Table 5 shows effect sizes which are adjusted to account for the differences in the time of measurement, and provide another index of program effectiveness.

TABLE 5: IMPROVEMENT ON VOCABULARY MEASURE AND EFFECT SIZES PER SCHOOL DURING THE 2007-2008 SCHOOL YEAR [5]

SCHOOL	PRE-TEST	POST-TEST	IMPROVEMENT/ MONTH	EFFECT SIZE
TREATMENT SCHOOLS				
REILLY	19.79	24.51	0.59	0.56
	(6.54)	(6.77)		
MERCER	18.01	22.02	0.50	0.40
	(6.14)	(7.15)		
WESTFIELD	16.85	20.55	0.46	0.33
	(6.29)	(7.39)		
MYSTIC	19.08	24.20	0.64	0.65
	(6.13)	(6.65)		
OCCIDENTAL	17.98	22.56	0.57	0.53
	(6.36)	(7.2)		
COMPARISON SCHOOLS				
GARFIELD	20.07	22.00	0.39	
	(6.48)	(7.3)		
JEFFERSON	20.85	21.97	0.22	
	(7.7)	(8.06)		
UXTON	21.67	24.47	0.56	
	(5.62)	(5.92)		
AVERAGES				
TREATMENT SCHOOLS	18.64	23.07	0.55	0.49
	(6.33)	(6.85)		
COMPARISON SCHOOLS	21.23	23.45	0.39	
	(6.38)	(6.85)		

[4] For a fuller description of these findings see Snow, C., Lawrence, J., & White, C. (2009).

[5] Adapted from Snow, C., Lawrence, J., & White, C. (2009).

PART III.
EVALUATING PROGRAM EFFECTIVENESS

Replication and Expansion in 2008-2009

In the second year, a new iteration of the curriculum was implemented, with 24 new topics and 120 new target words. The strategy for measuring student word learning during this year was the same as the previous year – a multiple-choice test with a selection of words from each week of the program, completed at the beginning and end of the year. Because we now had a history of working with the comparison schools, we were able in 2008 to administer the comparison pre-tests at the same time as in the treatment schools. We also modified the instructions to the teachers in ways designed to improve the student completion rates for the pre- and post-tests.

Both pre- and post-test data were available on 1183 students in seven treatment schools and 388 in three comparison schools. All students in the treatment schools received the intervention; those included in this analysis had completed usable test forms at both pre- and post-test. There were 810 girls and 770 boys in the analytic sample.

Assessment Challenges and Data Limitations

The team faced a different set of challenges in our efforts to evaluate program effectiveness in the 2008-2009 school year. Due to positive feedback on the program by principals in treatment schools, the second year of the quasi-experimental study saw an increase in the number of schools participating in the program, resulting in increased burdens on the program support staff. This increase unfortunately coincided with major financial difficulties in the district leading to announcements of school closings or restructurings (involving some schools that were implementing Word Generation). There were also high levels of absenteeism at the end of the year as a result of the H1N1 flu. Thus, there was considerable undertesting of students that was only partially offset by the improved instructions and oversight of the testing procedures.

These factors and others also resulted in less consistent program implementation in treatment schools, as demonstrated in the implementation analysis presented above.

TABLE 6: IMPROVEMENT ON VOCABULARY MEASURE AND EFFECT SIZES PER SCHOOL DURING THE 2008-2009 SCHOOL YEAR

	PRE-TEST	POST-TEST	IMPROVE-MENT	EFFECT SIZE	18 WEEK PRE-TEST	18 WEEK POST-TEST	IMPROVE-MENT	EFFECT SIZE
TREATMENT								
CARTER	18.69 (7.84)	20.44 (7.74)	1.75	-0.07	14.30 (6.02)	15.63 (5.94)	1.33	-0.02
LIPTON	16.82 (5.29	20.35 (6.20)	3.52	0.24	12.63 (3.99)	15.32 (4.83)	2.69	0.31
MERCER	18.24 (5.76)	20.68 (5.90)	2.45	0.03	13.86 (4.63)	15.85 (4.78)	1.98	0.12
MYSTIC	19.00 (5.33)	21.20 (5.56)	2.20	-0.01	14.33 (4.31)	15.98 (4.69)	1.65	0.05
OCCIDENTAL	13.40 (5.81)	17.16 (6.27)	3.76	0.26	10.32 (4.36)	12.72 (4.78)	2.39	0.22
REILLY	17.93 (6.10)	20.77 (6.35)	2.84	0.09	13.73 (4.72)	15.87 (5.08)	2.13	0.14
WESTFIELD	17.08 (5.35)	18.69 (5.88)	1.61	-0.12	13.01 (4.34)	14.08 (4.57)	1.07	-0.09
AVERAGE	**17.73 (5.97)**	**20.33 (6.21)**	**2.60**	**0.06**	**13.48 (4.69)**	**15.45 (4.97)**	**1.97**	**0.11**
COMPARISON								
JEFFERSON	19.22 (5.56)	20.93 (6.46)	1.70		14.57 (4.20)	15.66 (5.23)	1.09	
KENNEY	20.86 (4.52)	23.59 (4.61)	2.73		16.33 (3.51)	17.67 (4.21)	1.35	
UXTON	18.86 (5.87)	21.31 (5.50	2.45		14.42 (4.56)	16.06 (4.57)	1.64	
AVERAGE	**19.20 (5.67)**	**21.47 (5.74)**	**2.27**		**14.69 (4.38)**	**16.14 (4.75)**	**1.45**	

The results reflect the reduced fidelity and intensity of implementation. Table 6 shows pre- and post-test scores as well as effect sizes for all schools and by treatment and control conditions based on the 35 items that were in the curriculum, and differentiated for the 27 items that were taught in the first 18 weeks of the program. The effect sizes obtained from either calculation are lower than those obtained the previous year. The effect sizes obtained from analysis only of the items instructed in the first 18 weeks (Cohen's d = 0.11) are greater than those obtained based on analysis of all taught items (Cohen's d = 0.06), confirming that toward the end of the year implementation was increasingly uneven across school and content areas.

In addition to these implementation challenges, results indicate that the words chosen for the second year of the intervention were less challenging than those taught in year 1. Table 7 presents data from year 1 and year 2 in a common metric: the percentage of items scored correctly on the pre- and the post-test for each year. Notice that Word Generation students scored roughly five percent higher on the pre-test in year 2 compared to year 1, and comparison school students also seemed to find the year 2 words easier. The smaller effect sizes in 2008-2009 may reflect the fact that some of the words were too easy. Comparing absolute improvement levels suggests that the differences between the effect sizes during the two years was not the result of improved vocabulary instruction in the comparison schools, but rather a reduced impact in the Word Generation schools.

TABLE 7: IMPROVEMENT DURING 2007-2008 AND 2008-2009 EXPRESSED AS PERCENTAGES

	YEAR 1		YEAR 2		YEAR 1	YEAR 2
SCHOOL	PRE-TEST	POST-TEST	PRE-TEST	POST-TEST	IMPROVEMENT	IMPROVEMENT
TREATMENT						
CARTER			0.55 (0.23)	0.60 (0.23)		5.11%
LIPTON			0.49 (0.15)	0.59 (0.169)		10.34
MERCER	0.45 (0.15)	0.55 (0.18)	0.53 (0.18)	0.61 (0.18)	10.03%	7.63%
MYSTIC	0.48 (0.15)	0.60 (0.17)	0.55 (0.17)	0.61 (0.18)	12.89%	6.34%
OCCIDENTAL	0.45 (0.16)	0.56 (0.18)	0.40 (0.17)	0.49 (0.18)	11.45%	9.20%
REILLY	0.49 (0.16)	0.61 (0.17)	0.53 (0.18)	0.61 (0.20)	11.81%	8.20%
WESTFIELD	0.42 (0.16)	0.51 (0.18)	0.50 (0.17)	0.54 (0.18)	9.25%	4.12%
AVERAGE	**0.47 (0.16)**	**0.58 (0.18)**	**0.50 (0.18)**	**0.54 (0.19)**	**11.08%**	**7.59%**
COMPARISON						
GARFIELD	0.50 (0.16)	0.55 (0.18)			4.82%	
JEFFERSON	0.52 (0.19)	0.55 (0.20)	0.56 (0.16)	0.60 (0.20)	2.78%	4.19%
KENNEY			0.63 (0.13)	0.68 (0.16)		5.18%
UXTON	0.54 (0.14)	0.61 (0.15)	0.55 (0.18)	0.62 (0.18)	7.01%	6.29%
AVERAGE	**0.53 (0.16)**	**0.59 (0.17)**	**0.57 (0.17)**	**0.62 (0.18)**	**5.56%**	**5.56%**

PART III.
EVALUATING PROGRAM EFFECTIVENESS

EXPLORING LONG-TERM IMPACTS FOR DIFFERENT STUDENT GROUPS

Despite the evidence of vocabulary gains for Word Generation participants, we did not know if these gains were meaningful. Do students maintain knowledge of the words they have learned through summer vacation and the following school year? The goal of Word Generation is to improve vocabulary so that it results in improved reading comprehension. Clearly, short-term vocabulary learning will not generate long-term comprehension improvement. So to address this interest in the longer-term impact of the program, we conducted a follow-up longitudinal study to examine the effects of Word Generation on the learning, maintenance, and consolidation of academic vocabulary for students from English-speaking homes (EH), proficient English speakers from language-minority homes (LM not LEP), and limited English-proficient students (LEP). The results summarized here are detailed in a paper that is currently under review.[6]

Methods

As described previously, students in both the treatment and comparison schools completed a pre- and post-test on their knowledge of 48 of the instructed target words in the fall of 2007 and the spring of 2008. Similarly, pre-test and post-test data collected in fall 2008 and spring 2009 were designed primarily to assess the effectiveness of the 2008-2009 Word Generation implementation, so the majority of tested words had been instructed that year.

However, 11 items taken from the previous year's test were embedded in the 2008-2009 pre- and post-test, enabling us to conduct longitudinal analyses to determine if words learned were also maintained.

In order to construct a longitudinally-consistent measure and maximize the amount of information from the 11 items that were tested four times over two years, we used an item response theory (IRT) approach. First, we fit a single-factor model to the 11 items in each wave in order to test the hypothesis that the 11 items were reasonable indicators of a single factor of vocabulary knowledge (Muthén & Muthén, 2007). Then, we used the item parameters from wave one to produce scaled scores for each of the subsequent waves.

Longitudinal analytical methods allow for the flexible use of data (Singer & Willett, 2003). This flexibility allowed us to include all students who contributed at least one wave of data during 2007-2008 in our analysis, although we did not include students who only contributed data during the third (fall 2008) or fourth (spring 2009) waves since we could not be sure that these students had received instruction on the target words and we were particularly worried about the high mobility rates of our limited-English proficiency (LEP) students. This process resulted in no cases being dropped for the first two waves of data but the exclusion of many students who entered the study during the second year.

[6] Lawrence, J., Capotosto, L., Branum-Martin, L., White, C., & Snow, C. (in revision). *Language proficiency, home-language status, and English vocabulary development: A longitudinal follow-up of the Word Generation program.*

TABLE 8: LONGITUDINAL PERFORMANCE IN TREATMENT AND COMPARISON SCHOOLS[7]

SCHOOLS		SCALED				RAW			
		INSTRUCTIONAL YEAR		FOLLOW UP YEAR		INSTRUCTIONAL YEAR		FOLLOW UP YEAR	
		FALL 2007	SPRING 2008	FALL 2008	SPRING 2009	FALL 2007	SPRING 2008	FALL 2008	SPRING 2009
TREATMENT									
Reilly	Mean	-0.088	0.473	0.108	0.442	4.666	6.099	5.682	6.719
	SD	(0.728)	(0.793)	(0.772)	(0.826)	(2.122)	(2.172)	(2.306)	(2.460)
	N	329	382	223	210	329	382	223	210
Mercer	Mean	-0.047	0.445	0.098	0.487	4.835	5.841	5.674	6.562
	SD	(0.752)	(0.859)	(0.835)	(0.795)	(2.194)	(2.443)	(2.448)	(2.271)
	N	468	391	279	267				
Westfield	Mean	-0.215	0.195	-0.193	0.262	4.254	5.116	4.679	5.971
	SD	(0.672)	(0.786)	(0.832)	(0.864)	(2.069)	(2.227)	(2.422)	(2.671)
	N	114	155	109	68	114	155	109	68
Mystic	Mean	-0.017	0.559	0.150	0.491	4.883	6.134	5.687	6.670
	SD	(0.705)	(0.803)	(0.712)	(0.873)	(2.019)	(2.192)	(2.044)	(2.478)
	N	137	149	99	97	137	149	99	97
Occidental	Mean	-0.305	0.214	-0.355	0.132	4.087	5.431	4.064	5.684
	SD	(0.672)	(0.890)	(0.639)	(0.803)	(1.948)	(2.507)	(2.151)	(2.395)
	N	92	102	47	38	92	102	47	38
Average	**Mean**	**-0.093**	**0.416**	**0.038**	**0.431**	**4.674**	**5.831**	**5.435**	**6.518**
	SD	**(0.765)**	**(0.872)**	**(0.780)**	**(0.856)**	**(2.132)**	**(2.326)**	**(2.382)**	**(2.419)**
	N	**1140**	**1179**	**757**	**680**	**1140**	**1179**	**757**	**680**
COMPARISON									
Walters	Mean	0.227	n.a.	n.a.	n.a.	5.696	n.a.	n.a.	n.a.
	SD	(0.687)	n.a.	n.a.	n.a.	(2.031)	n.a.	n.a.	n.a.
	N	92	0	0	0		0	0	0
Garfield	Mean	0.096	0.396	n.a.	0.308	5.375	5.754	n.a.	6.200
	SD	(0.772)	(0.860)	n.a.	(0.788)	(2.293)	(2.340)	n.a.	(2.151)
	N	56	57	0	40	56	57	0	40
Jefferson	Mean	0.089	0.348	-0.036	0.245	5.205	5.412	5.236	5.887
	SD	(0.848)	(0.927)	(0.763)	(0.945)	(2.398)	(2.592)	(2.359)	(2.729)
	N	112	119	72	62	112	119	72	62
Uxton	Mean	0.250	0.666	0.254	0.718	5.747	6.493	6.061	7.213
	SD	(0.751)	(0.826)	(0.775)	(0.792)	(2.174)	(2.212)	(2.269)	(2.245)
	N	265	229	131	155				
Average	**Mean**	**0.195**	**0.534**	**0.150**	**0.540**	**5.583**	**6.072**	**5.765**	**6.735**
	SD	**(0.729)**	**(0.831)**	**(0.802)**	**(0.827)**	**(2.218)**	**(2.393)**	**(2.324)**	**(2.422)**
	N	**525**	**405**	**204**	**257**	**525**	**405**	**204**	**257**

[7] Lawrence, J., Capotosto, L., Branum-Martin, L., White, C., & Snow, C. (in revision). *Language proficiency, home-language status, and English vocabulary development: A longitudinal follow-up of the Word Generation program.*

PART III.
EVALUATING PROGRAM EFFECTIVENESS

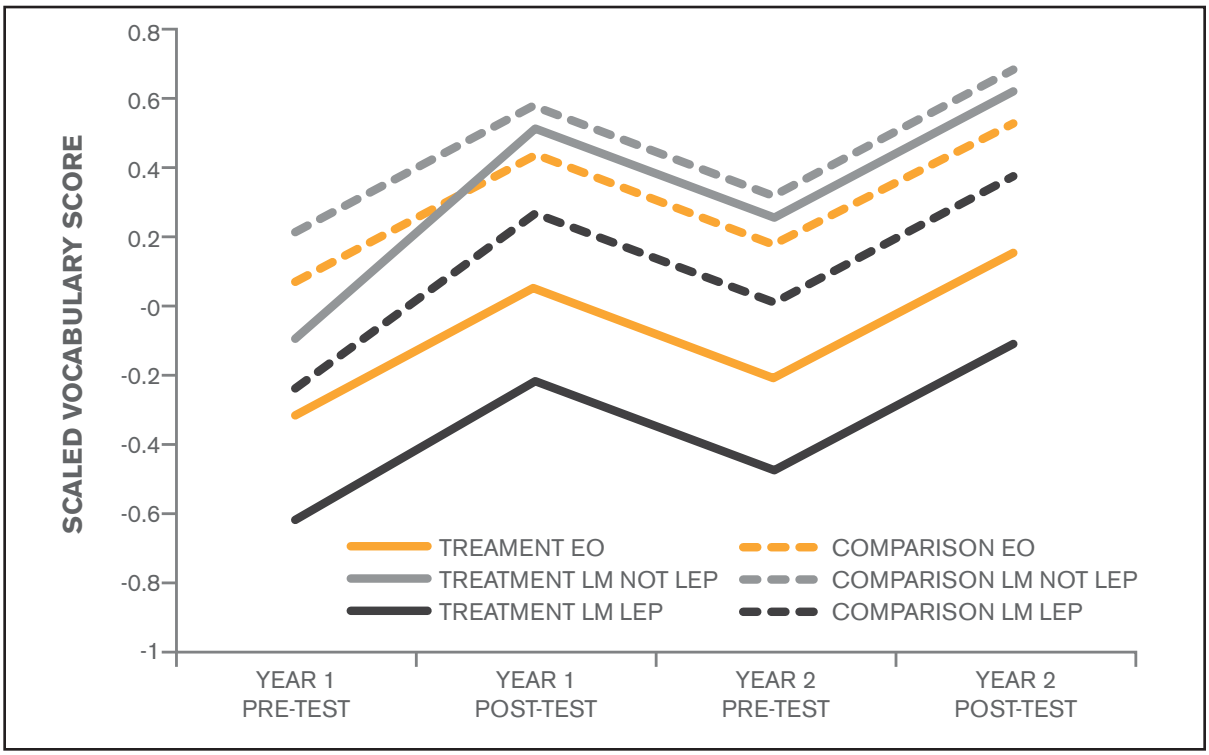
Findings

Table 8 presents raw and scaled vocabulary data from comparison and treatment school students across the four waves of data. To illuminate subgroup differences, Figure 3 presents these data separately for English-only students, Language Minority students who are English Proficient, and Limited English Proficient students. In this figure, dotted lines represent students from comparison schools, and solid lines represent students who had received the Word Generation curriculum. At baseline (fall 2007), comparison school students in all home-language and language-proficiency categories scored better on vocabulary knowledge than their treatment peers. In both groups, surprisingly, English-proficient students from language minority homes began the study with somewhat stronger vocabulary knowledge than English-proficient students from English-speaking homes. Differences between proficient and LEP students were pronounced at all four waves of data collection for both treatment and comparison school students.

We used growth modeling techniques to determine how much English-proficient students from LM versus EO homes benefited from program participation, and how well they maintained vocabulary knowledge during summer and the following school year. As can be seen from Figure 3, treatment students made stronger gains than students in the comparison schools during the intervention period – as shown by the steeper slopes of the lines representing those groups between points 1 and 2 on the horizontal axis. Furthermore, gains were larger for language minority students than for students from English-speaking homes; not only does the line indicating their growth rise steeply, but it even crosses the line for English-only students in comparison schools.[8]

The current study allowed us to pinpoint the long-term effect of program participation on student vocabulary for EO, LM, and LEP students. English-proficient students from language-minority homes who participated in the program made strong gains – gains that put their scores above those of EO students in comparison schools –

FIGURE 3: PROTOTYPICAL PERFORMANCE OF 6TH GRADE STUDENTS IN TREATMENT AND COMPARISON GROUPS, COMPARING ENGLISH ONLY STUDENTS, ENGLISH PROFICIENT STUDENTS FROM LANGUAGE MINORITY HOMES, AND STUDENTS OF LIMITED ENGLISH PROFICIENCY.



[8] See Snow, C., Lawrence, J., and White, C. (2009). for a different analysis also indicating that language minority students gained more from the Word Generation curriculum than English-only students.

from the intervention. Furthermore, they maintained those gains relative to comparison students even a year later. English proficient students from English-speaking homes also made gains relative to the comparison group and maintained them across the course of the study. However, LEP students did not show comparative benefits from participation in the Word Generation program; their rate of growth continued to parallel that of their LEP peers in the higher achieving schools, with no narrowing of the gap.

EXAMINING THE RELATIONSHIP OF WORD GENERATION PARTICIPATION TO MCAS SCORES

In the absence of a proper experimental study, we are unable to make strong inferences about the impact of the Word Generation program on external measures, such as the Massachusetts Comprehension Assessment System (MCAS). However, we conducted an exploratory analysis to determine whether the number of Word Generation words a student learned was associated with MCAS scores from the end of that academic year.[9]

We had already determined, in previous analyses, that students who scored under 80% correct on our pretest were unlikely to have performed in the proficient or advanced range on the MCAS, but of course that finding simply confirms the importance of academic vocabulary as a predictor of test outcomes. The analysis of interest was designed to determine whether growth in academic word knowledge predicted MCAS scores better for students participating in Word Generation than for those not participating. If there was a difference, that would support the claim that participation in Word Generation constituted good preparation for MCAS.

In order to determine whether there was a relationship between participation in Word Generation and performance on the MCAS, we performed regression analysis using a model with gender, treatment status, pretest and posttest scores as predictors of April 2008 MCAS scores. The addition of an interaction term also allowed us to measure whether post test scores interacted with treatment in predicting MCAS scores (controlling for pretest scores).

For a full description of the methodology and findings, see Snow, Lawrence, and White, 2009. In summary, our results indicated that this interaction was significant and improved the model, which suggested to us that it wasn't just vocabulary knowledge or program participation alone, but the interaction of both elements – evidence of vocabulary development and participation in the Word Generation program – that improved our ability to predict MCAS scores.

Our next step was to further examine this interaction between treatment and vocabulary improvement by creating separate models to predict MCAS scores for the treatment and comparison schools. We found that the model created for Word Generation schools predicted MCAS achievement better than the model created for comparison schools, and in the Word Generation schools student post-test scores were much stronger predictors of MCAS achievement than pre-test scores. Again, this suggests that post-test scores in Word Generation schools captured not only target vocabulary knowledge at the end of the year, but also student participation level in the Word Generation program.

Of course, these analyses do not control for baseline reading achievement scores, which were available for some but not all of the students in our sample. Nor do they account for 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 (percent LM students) of those schools. These are important limitations to keep in mind in interpreting the findings of these early evaluation efforts, and point to important directions for future work and research.

[9] See Snow, C., Lawrence, J., and White, C. (2009).

DISCUSSION

DISCUSSION

IMPLICATIONS OF FINDINGS FROM QUASI-EXPERIMENTAL STUDY

The findings of this quasi-experimental study were highly informative, both about the potential of the Word Generation approach to support students' academic progress, and about the challenges to an optimal implementation of the program. We were highly gratified to see strong and lasting vocabulary advances for students from language minority homes, precisely because these may be disadvantaged in the domain of academic, school-related vocabulary by lack of exposure to it at home. We were also encouraged that the gains made by those and by English-only students as a result of participation in the program were maintained and even enhanced during the following school year. We hypothesize that the post-program gain in knowledge of academic words reflects recurrent exposure to those words, because they appear in texts students read and in classroom discourse.

It is also encouraging that post-test scores on the Word Generation assessments strongly related to performance on the state accountability assessment. It seems obvious that this may simply reflect the use of the taught words on the state test. However, this explanation is undermined by the absence of a similarly strong relationship in the treatment schools. Furthermore, while improvement in the Word Generation schools was significant, it was still modest – about four words out of forty 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 the post test achievement on the multiple choice 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. Since the Massachusetts test is a relatively challenging one (arguably the best aligned with the NAEP 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.

The disappointing outcomes for LEP students, on the other hand, may be explained by the challenge of the program, or perhaps by their lack of access to classroom activities of the level necessary to reinforce the effect of the program. Certainly Word Generation would not be advised for beginning-level LEP students; they need to master basic English before embarking on the content or the language of this program. But we feel that access to the topics and the activities embedded in Word Generation is crucial for somewhat more advanced LEP students, if they are to make the transition to regular classroom work. We are thus seeking opportunities to test adaptations of Word Generation to ESL and bilingual education settings.

ONGOING WORK

The work reported here represents the early stages of our efforts to evaluate the effectiveness of Word Generation. It provided sufficient empirical indication of feasibility that a proposal to IES to fund a proper experimental study was honored. The experiment is currently being conducted in the Baltimore City Schools, Pittsburgh Public Schools, and San Francisco Unified School District. In addition, a version of Word Generation designed to be implemented four days a week in English language arts has been developed and tested in a couple of Austin, TX schools, under the auspices of CREATE (http://www.cal.org/create/), and in conjunction with parallel interventions focused on science and social studies. The all-ELA Word Generation retains the passage discussion, debate, and writing components, and adds in a set of word study activities, using target words as a launch for teaching morphological analysis, cognate use, and common root words.

Word Generation also forms a centerpiece of work recently funded by IES under the Reading for Understanding Initiative. SERP and Harvard University have been funded to study ways to enhance reading comprehension among students in grades 4-8. We proposed to extend Word Generation downwards to grades 4 and 5, and to enhance Word Generation across

the grades by developing some extended units focused on particular topics, rather than shifting topic every week. The extended topics are designed provide the opportunity for students to accumulate more relevant background knowledge, and to work during an extended period on a longer piece of writing.

While the research activities around Word Generation continue, practitioners are embracing the program even in advance of experimental findings. Registrations on www.wordgeneration.org to download the program materials numbered above 3000 by October, 2010. While we assume that many people download the program out of curiosity rather than with the intention of implementing it, we know from email and other feedback that many have tried it. The Boston schools that were early and enthusiastic implementers have been visited by delegations from other districts interested in adopting Word Generation, and from schools as far away as Norway and The Netherlands. Practitioners also suggest improvements to the program, and modify it to their own purposes. Patrick Hurley, for example, who teaches at Mountain View High School in California, has adapted and enhanced the program for use with his high school ESL students (see Hurley, 2010).

REFLECTIONS ON WORKING COLLABORATIVELY

Word Generation has been a product of the SERP commitment to collaboration between practitioners and researchers. Both groups have made important contributions to the ongoing work and to the final product. The researchers have insisted on embedding in the program features reflecting what we know from studies of effective vocabulary teaching, and on collecting data to inform schools implementing the program and those interested in doing so about features of effective implementation and about impact. The practitioners provided the initial impetus to focus on all-purpose academic vocabulary, and offered ongoing feedback about the appropriateness of topics chosen for the weekly dilemmas, about the right challenge level of the activities and problems, about what kind of professional development and support they needed to implement the program effectively, and about how to align the program activities with district priorities. These lessons are being put to good use in our current Reading for Understanding grant activities, as we extend Word Generation down to grades 4 and 5, and develop more extended reading, writing, and discussion activities linked to district standards for science and for social studies.

It would be naïve to suggest that the collaborative efforts around Word Generation have all gone smoothly. Some of the schools involved struggled to organize the sequence of activities and the availability of the student wordbooks at the right times in the right places. There were reluctant participants in some schools, and even when implementation was consistent it was by no means universally excellent. What the researchers and program developers intended as a resource for the schools was sometimes seen as a burden by the teachers who were using it.

Furthermore, each year the process of recruiting schools, scheduling professional development and pre-test sessions, and distributing materials runs into new snags. Indeed, without the stable presence of the SERP partnership, the access to District central office personnel the partnership structure provides, and the history of mutual commitments as a foundation for this work, it would likely have foundered several times over the last few years. In fact, the Word Generation program in 2010-2011 is being fully implemented in only three BPS schools. The work done in Boston, though, in the context of the SERP partnership, has attracted wide attention, with the result that dozens of schools and teachers across the country (and internationally) are using the Word Generation materials and implementing the Word Generation model, in which active discussion about engaging topics invites students into the use of sophisticated, academic language.

ACKNOWLEDGEMENTS

The authors would like to thank our collaborators including Claire White (Strategic Educational Research Partnership), Lauren Capotosto (Harvard University) and Lee Branum-Martin (University of Houston) for their collaboration on previous research which has been synthesized in this report.

REFERENCES

REFERENCES

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

Donovan, M. S., Wigdor, A.K., & Snow, C.E. (Editors). (2003). Strategic Education Research Partnership. Washington, DC: National Academies Press.

Fuhrman, S., & Elmore, R. F. (2004). Redesigning accountability systems for education. New York: Teachers College Press.

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

Hurley, P. (2010). Academic language: Equipping English learners to speak and write confidently in secondary classrooms. Gifted Education Communicator, Winter, 21-25.

Lawrence, J., Capotosto, L., Branum-Martin, L., White, C., & Snow, C. (under review). Learning and maintaining academic vocabulary.

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.

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

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 (No. NIH Publication No. 00-4769). Washington, DC: U.S.: U.S. Government Printing Office.

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

Snow, C., Lawrence, J., & White, C. (2009). Generating knowledge of academic language among urban middle school students. Journal of Research on Educational Effectiveness, 2(4), 325-344.

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, New Jersey: Lawrence Erlbaum Associates.

THE COUNCIL OF THE GREAT CITY SCHOOLS

1301 Pennsylvania Avenue, NW
Suite 702
Washington, DC 20004

202-393-2427
202-393-2400 (fax)
www.cgcs.org