Generating Knowledge of Academic Language among Urban Middle School Students

Catherine E. Snow, Joshua F. Lawrence, Claire White

Background/context. The reading comprehension of post-primary grade students, in particular those attending urban schools, is a matter of recurrent concern. Performance of 8th graders on the NAEP reading assessment, for example, shows that 74% of all students perform at or above the basic (grade-appropriate) level, whereas only 60% of students in the large central urban districts perform at that level. Ethnic and language minority students in these districts perform even less well than Caucasian and native English-speaking students. While the specific reading challenges faced by students performing below basic level are no doubt heterogeneous, a ubiquitous issue mentioned by their teachers and confirmed by assessment is their limited vocabularies. It is not surprising that the many language minority students in urban districts show gaps in English vocabulary, but even native English speakers may fail to develop rich vocabularies if they have a history of low reading ability, limited comprehension, and low investment of time in reading, because much sophisticated vocabulary is acquired through reading (Anderson, Wilson, & Fielding, 1988; Stanovich, 1986).

Purpose / objective / research question / focus of study. Our purpose was to understand: 1) how well students participating in a vocabulary program learned target words relative to other students; 2) if treatment effects were better for language minority (LM) or English only (EO) students, and; 3) if improved vocabulary predicted improved scores on a state-mandated standardized assessment.

Setting. Five treatment and three comparison middle schools in Boston Public Schools (BPS) in Massachusetts. Roughly half the students in the schools were LM students.

Population / Participants / Subjects. This is a quasi-experimental study in which academic word-learning by students in five schools implementing the Word Generation program was compared to academic word-learning by students in three schools within the same system that did not choose to implement the program. Because the implementing schools were those that volunteered for the program, selection effects must be taken into account in interpreting the findings.

Participants and Setting

  Schools. Word Generation was implemented in five schools during the 2007-2008 academic year, three middle schools and two K-8 schools in which only the 6th-8th grades used the program. Two schools, the Reilley and the Westfield, were completing their second year of implementation in 2007-2008, while the Mystic, Occidental, and Mercer Schools launched Word Generation in fall 2007 (pseudonyms are used for all schools). Demographics of the Word Generation and comparison schools reflect BPS more broadly, with a high incidence of poverty (ranging from a low of 79% to a high of 91% students receiving free or reduced-price lunch). BPS is characterized in general by rather high levels of special education designation, and all the schools shared this feature (between 16% and 33% of students with IEPs). A very high proportion of students at these schools come from second language homes, with percentages ranging from 32% to 70% across the schools. Four of the treatment schools offered Sheltered English Immersion (SEI) services to their limited English proficient (LEP) students; all students

enrolled in these sheltered classrooms (who represented between 6% and 26% of their school populations) received the Word Generation curriculum, albeit with modifications such as extended time and translation of key concepts.

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

The five implementing schools participated in professional development activities to varying degrees, because of difficulties scheduling and organizing the required meetings. For example, the Mercer received only one brief PD session, whereas the Occidental participated in a four-day summer institute, received eight hours of PD prior to launch, and engaged in biweekly cross-grade school-site sessions devoted to feedback on and previewing of the materials and activities, with support from the Word Generation team at several of those sessions.

Students. Both pre- and post-test data were available on 697 6th, 7th, and 8th grade students in five treatment schools and 319 in three comparison schools. All students in the treatment schools received the intervention; those represented in this data set had completed usable test forms at both pre- and post-test. There were 349 girls and 348 boys in the treatment schools, and 162 girls and 157 boys in comparison schools. Of these, 438 were classified as LM (parents reported preferring to receive materials in a language other than English), 287 in treatment schools and 151 in comparison schools. The vast majority of students in both treatment and comparison schools were low-income.

Intervention / Program / Practice. Word Generation is a 24-week-long sequence of topics of current interest, each associated with five all-purpose academic words, and prescribed activities related to math, science, and social studies. The basic sequence of Word Generation activities was the following: On Monday a brief text in which the five target words were embedded was read by the students and teacher together, then discussed using guiding comprehension questions; this text presented arguments on both sides of some difficult controversy or dilemma. Then the five target words were highlighted and provided with student-friendly context-related definitions. This activity typically occurred in the English Language Arts classroom. On Tuesday, Wednesday, and Thursday, in an order determined by each school, the math, social studies, and science teachers respectively implemented activities provided for them, each of which embedded the same five target words. The math teacher assigned one or two problems related in content to the dilemma of the week; the format of these problems was modelled on the state math assessment. Math teachers then discussed the content as well as the math procedures. The science teacher presented a new text that focused on science content related to the dilemma of the week; students filled in target words left blank in the text, before the class discussed the text. The social studies teacher organized a debate about the dilemma of the week in one of several possible formats (fishbowl, pairs, whole class, four corners, etc.). On Friday, the students were asked to write a 'taking a stand' essay about the dilemma.

Various aspects of the Word Generation design respond to the local conditions in the district for which it was originally developed. Most 6th-8th graders in BPS attend separate middle schools where content area instruction is departmentalized, and teacher planning time built into the school schedule typically occurs within departments, limiting the opportunities for teachers to share information about student progress or curricular emphases across those departmental

boundaries. Transcending the boundaries to recruit participation by all the teachers in vocabulary teaching was one goal of the Word Generation design.

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

Research Design: This was a quasi-experimental evaluation of an intervention implemented at the school level. Five schools participated in the intervention and there were three comparison schools. Effects sizes were calculated by comparing the improvement students made from pre to post on a vocabulary assessment in treatment schools with improvement made in comparison schools. Secondary analyses examined language status as a moderator of word learning and word learning as a mediator of improved reading comprehension on a state mandated English language arts assessment.

Data Collection and Analysis: The efficacy of the intervention was assessed using a 48-item multiple choice test that randomly sampled two of the five words taught each week. The vocabulary assessment was not completed by all students in the time available. Because items at the end of the assessment had particularly low rates of completion, we dropped the last four items from our analysis of both pre- and post-test. The reliability of the test with the 40 items that remained was acceptable (Cronbach's alpha = .876).

This instrument was administered to students in all the treatment schools in October 2007, before the introduction of Word Generation materials. Because of difficulty recruiting the comparison schools, the pretest was not administered there until January. The 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 analyses in terms of words learned per month as well as total words learned.

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

Findings / Results: Descriptive statistics show that students in the Word Generation program learned approximately the number of words that differentiated 8th from 6th graders on the pretest—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 pretest. In order to account for the differences in test administration times, the pre to post improvement in all schools was divided by the number of months between the pre and post test administration: the average improvement per months in the treatment schools was greater than that in the comparison schools. The average effect size on the researcher developed vocabulary assessment in the treatments schools was 0.49 (controlling for the improvement ascertained in the comparison schools).

Regression analysis was used to determine if participation in Word Generation predicted improved vocabulary outcomes, controlling for the pretest. Gender was a significant predictor of word learning ($\beta = -0.052$, $p < 0.007$), as was treatment ($\beta = 0.166$, $p < 0.001$). Language status (LM versus EO) was not a significant predictor, but the interaction of treatment and language status was at the margin of significance ($p = 0.055$), and including the interaction improved the overall model. Interestingly, student pretest vocabulary did not interact with treatment in predicting posttest scores. We split the data set to investigate the home language variable more closely. The first set of regressions used pretests and gender to predict post-test scores in the comparison schools (r squared = 0.62) and Word Generation schools (r squared = 0.64). In Word Generation schools LM status predicted improved vocabulary ($\beta = -0.053$, $p = .022$), but it was not a significant predictor in comparison schools.

In order to determine whether participation in Word Generation had any relationship to performance on the MCAS, a regression model was fit with MCAS scores in April, 2008 as the outcome, using gender, treatment status, pre-test and post-test scores as predictors. We added an interaction term to see if post-test scores interacted with treatment in predicting MCAS scores (controlling for pretest scores). The interaction term was significant ($\beta = .21$, $p = 0.01$) and its inclusion improved the model.

We further explored the interaction between treatment and vocabulary improvement by splitting the data and refitting the models to data from the treatment and comparison school separately. The fitted model for comparison school data did not predict MCAS achievement (R Square = 0.41) as well as the fitted model for the treatment school data (R Square = 0.49). In the Word Generation schools student post-test scores ($\beta = 0.527$, $p < .001$) were much stronger predictors of MCAS achievement than pre-test scores were ($\beta = 0.201$, $p < .001$), perhaps because the post-test scores captured not only target vocabulary knowledge at the end of the year, but also level of student participation in the Word Generation program.

Unfortunately, these analyses do not control for baseline reading achievement data, which were available only for a subset of students in our sample (n = 389). For that subgroup, we used fall standardized reading comprehension scores (on the GRADE) as a covariate to determine if the relation between improved vocabulary and MCAS persisted even when controlling for overall reading levels. Results demonstrate both that the GRADE is a strong predictor of spring MCAS scores ($\beta = 0.750$, $p < .001$) and that the interaction between treatment and improvement persists in the model controlling for GRADE. Split file analysis demonstrated the familiar pattern, with vocabulary improvement predicting MCAS scores for student in the treatment schools ($\beta = 0.151$, $p < .001$) but not for students in the comparison schools. GRADE scores were also used to determine if better readers learned words more efficiently than less able readers. Results demonstrate that GRADE baseline scores did not predict word learning and that there was no significant interaction between treatment and baseline reading achievement as measured on the GRADE.

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

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

It is particularly encouraging that post test scores on the Word Generation assessments strongly related to performance on the state accountability assessment. One might assume this reflects the coincidence that the words taught also occurred on the state test. However, this simple explanation is undermined by the absence of a similarly strong relationship in the treatment schools. Furthermore, 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 improvement on our curriculum-based assessment represents an index of exposure to the Word Generation curriculum – a curriculum that taught new content, deep reading and comprehension skills, discussion, argumentation, and writing. 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.

Appendices

Appendix A References

Carlo, M., August, D., McLaughlin, B., Snow, C., Dressler, C., Lippman, D., et al. (2004).

Closing the gap: Addressing the vocabulary needs of English-language learners in

bilingual and mainstream classrooms. *Reading Research Quarterly, 39*(2), 188-215.

McBeath, J., Reyes, M.E., Ehrlander, M.F. (2007). *Education reform in the American states.*

Charlotte, NC: IAP.

Stahl, S., & Fairbanks, M. (1986). The effects of vocabulary instruction: A model-based meta-

analysis. *Review of Educational Research, 56*(1), 72-110.