Abstract 2 Title Page

Title: Experimental Effects of Word Generation on Reading Performance in High Poverty Middle Schools

Authors and Affiliations:

Stephanie M. Jones, Harvard Graduate School of Education
James Kim, Harvard Graduate School of Education
Maria LaRusso, Harvard Graduate School of Education
Ha Yeon Kim, Harvard Graduate School of Education
Catherine Snow, Harvard Graduate School of Education

Abstract 2 Body

Background / Context:
Time to Act, a 2009 report of the Carnegie Corporation's Council on Advancing Adolescent Literacy, concludes that U.S. students are ill-prepared for the literacy challenges of 21st century higher education, employment, and citizenship. The poor performance of U.S. high schoolers in international comparisons contrasts sharply with the relatively good performance of U.S. 4th graders. The success of 4th graders, and indeed the recent rise in 4th grade NAEP scores, is believed to reflect the success of federal and state policies focused on primary literacy. It seems we have learned to teach students to read at the 4th grade level without preparing them for subsequent literacy challenges. As a result, a high proportion of middle and high school students are struggling. These strugglers are overrepresented in urban districts, among students living below the poverty line, and among ethnic and linguistic minorities.

Improving reading for understanding in the post-primary grades requires exposing students who read at all levels to new instructional elements that focus on higher-order comprehension skills (e.g., analysis, synthesis, critique, problem-solving). To target these higher order skills, teachers need (1) a better understanding of the component skills required and how they develop, (2) a set of digestible instructional activities that, if well executed, build these skills, and (3) opportunities to learn the classroom discussion procedures that support student progress. High quality discussions are critical to three basic components of reading comprehension: perspective-taking, complex reasoning, and academic language skill. For example, many children have the opportunity to develop academic language skills in non-academic contexts, for example, in dinner-table conversations with engaged adults or while reading for pleasure or information. For many others, though, access to such skills is limited to the classroom—they are entirely dependent on the richness of the school language environment to support their acquisition of academic language. Furthermore, if they are poor or reluctant readers then they are almost solely dependent on the *oral* language environment of the classroom for their exposure to academic language.

Word Generation (WG) is a research-based vocabulary program for middle school students designed to teach words through language arts, math, science, and social studies classes. The program consists of weekly units that introduce 5 high-utility target words through brief passages designed to spark active examination and discussion of contemporary issues. WG was designed with the understanding that promoting classroom discussion can result in particular kinds of academic benefits, such as improved word knowledge, reasoning, and expression. The IES funded evaluation of WG (as part of the Reading for Understanding initiative) is a school-level experimental study that includes two cohorts of schools randomized to treatment and control conditions. The first cohort has been studied for three years and the second cohort for two years, although in this presentation we focus only on the first two years of the study as a whole.

Purpose / Objective / Research Question / Focus of Study:
In this paper we present new data from the first two years of the WG evaluation focusing on:

1. basic psychometric and descriptive information (including variance decomposition), by grade level, for measures of academic language and perspective taking as well as

standardized assessments of student vocabulary and reading comprehension (e.g., *Word Generation Academic Vocabulary Assessment and Global Integrated Scenario-based Assessments, GISA*); relationships among these measures and the standardized assessments within and across time, and by grade level; and

2. results from impact analyses focusing on WG intent-to-treat effects on (a) standardized assessments of student vocabulary and reading comprehension, and (b) new measures of academic language and perspective taking after one and then two years for the first cohort of schools, and after one year for the second cohort of schools.

Setting:
As part of the IES-funded project *Catalyzing Comprehension through Discussion and Debate* (CCDD), the data for this study were collected as part of the impact evaluation of Word Generation. In this paper we focus on 4th graders in cohort 1 (6 schools) in year 1, following those children into 5th grade in year 2. With the addition of cohort 2 in year 2, we focus on 4th and 5th graders from an additional 18 schools., the two cohorts represent four school districts. The districts, located in Massachusetts and Maryland, include two major cities serving ethnically diverse, low income students; one small city serving ethnically diverse, primarily low income students; and one suburban district serving a primarily white, low to middle income population.

Population / Participants / Subjects:
As noted above, the CCDD evaluation sample is socio-demographically diverse. For example, in district 1, percentages of students in the participating schools scoring below proficient on the 2013 state ELA assessment ranged from 55 to 79, between 79 and 91 percent of the student population was African American or Hispanic, and between 83 and 92 percent were eligible for free or reduced price lunch. In district 4, 38% of the students scored below proficient in ELA. Seventeen percent were African American or Hispanic and 44 percent were subsidized lunch eligible.

Intervention / Program / Practice:
Word Generation (WG) is a research-based vocabulary program for middle school students designed to teach words through language arts, math, science, and social studies classes. The program consists of weekly units that introduce 5 high-utility target words through brief passages designed to spark active examination and discussion of contemporary issues. WG was designed with the understanding that promoting classroom discussion can result in particular academic benefits, such as improved word knowledge, complex reasoning, and perspective-taking. Each lesson is constructed around a text that provides a bit of information about a controversial issue, and a few arguments on either side of the issue. Some of the issues are local and student-centered, e.g., Should junk food be banned from school cafeterias? Should the school day be lengthened? Others are more national or global in scope, e.g., Should the government impose a mandatory year of service after high school? Should physician-assisted suicide be legal? The launch text introducing each issue is written in journalistic-academic language, and incorporated academic vocabulary words, five of which are then targeted for explicit instruction in the course of that week. To support students in using the target words, there is a debate on the issue of the week. Finally, drawing on the debate as well as the texts as a resource, the students are asked to write a 'taking a stand' paragraph at the end of the week, also using the target words when possible.

Research Design:
Twenty four schools in total were randomized within the four districts, within pairs of schools matched on several characteristics such as enrollment, socio-demographics, and standardized test scores. The program was implemented and data were collected on all students in grades 4 through 7. Data were collected for three years (2011-2014) with the final wave of data collection completed in June 2014. For this paper we employ data collected in the first two years of the study focusing on 4th graders in cohort 1 in year 1, following those children into 5th grade in year 2 (with two data collection points, fall and spring, in each year), and on 4th and 5th graders in cohort 2 in year 2 (again with two data collection points, fall and spring).

Data Collection and Analysis:
<u>Measures</u>
All students in the participating grades completed the following group-administered assessments at all time points noted above:

WG Academic Vocabulary. Multiple-choice test that assesses academic words targeted by the WG intervention in grades 4 to 8. The majority of target words are selected from the Academic Word List (Coxhead, 2000). Each target word is presented in a neutral sentence context with four responses to choose from: (a) a synonym (correct answer), an incorrect semantic associate, a phonological associate and a nonrelated word. The test includes 50 items. Percent correct scores were used for analysis (Hwang, Lawrence, & Snow, in preparation).

Global, Integrated Student Assessments (GISA). The Educational Testing Service was awarded a grant from the Department of Education as part of the same Institute of Education Sciences' Reading for Understanding Initiative to develop innovative reading comprehension assessments. The GISA is designed to be more engaging for students and more useful for teachers than previous assessments. These assessments use scenarios, technology, and reading strategies to motivate students, to model skilled reading, and to help disentangle key areas for improvement. These computer-based assessments are scenario-driven, providing a plausible purpose for reading that guides the assessment activities. Example scenarios include having students imagine that they are preparing to lead a class discussion or that they are part of a study group. All of the activities focus on strategic reading behaviors that skilled readers use every day. GISA batteries cover a range of content areas, including Science, Social Studies, and English Language Arts and a variety of text types that students regularly encounter (expository texts, fiction, e-mail, web pages, and blogs). Skills assessed by the GISA include: reading comprehension, perspective-taking, summary writing, metacognition, identifying topical vocabulary and learning vocabulary in context, questioning, identifying main ideas, sequencing, organizing key words and information, and paraphrasing.

Core Academic Language Skills Instrument (CALS-I). Group-administered test that evaluates students' core academic language skills in grades 4 to 8. The purpose of this test is to assess students' skills in understanding, producing, and reflecting upon language forms that are prevalent in academic texts (e.g., logical connectives, nominalizations). Tasks assess a range of skills through multiple choice, matching, or short written responses. Two statistically

equated forms with robust psychometric properties comprise the CALS-I: For Form 1 (for grades 4 to 6) and Form 2 (for grades 7 and 8). Form 1 was used for this study and included six tasks: Connecting Ideas, Tracking Themes, Organizing Texts, Breaking Words, Comprehending Sentences, Identifying Definitions. Prior studies have yielded robust reliability (.93 as indexed by coefficient alpha and .90 by split half reliability) and validity (.70 as indexed by the zero order correlation with the Gates-MacGinitie Passage Comprehension). Raw scores were used for this analysis, but IRT factors scores will be generated for the final analysis. Past studies have shown raw scores and IRT factors scores to be highly correlated.

The Perspective Taking Survey (PTS). Developed by Robert Selman and colleagues and designed to assess students' abilities to acknowledge, articulate, and position the perspectives of multiple stakeholders in a given text presenting a social conflict, and provide solutions that consider and integrate their different positions. The PTS includes three scenarios and each scenario is followed by 9 open-ended questions that are scored using a coding manual, achieving good inter-rater reliability among different coders.

Findings / Results:
Preliminary data from the first year (cohort 1) of the WG evaluation indicate that, as expected, mean levels of the primary constructs of focus increase significantly over the course of one school year, and are significantly higher for older grades than for younger (e.g., 4th graders). In addition, the measures of academic language and perspective taking were correlated with standard assessments (e.g., with GISA, ~.5-.6), and the intercorrelations did not vary substantially by grade.

Preliminary impact analyses using ordinary least squares (OLS) regression models from the first year (cohort 1) of the WG evaluation, focusing on the WG Academic Vocabulary assessment, revealed a positive impact of WG on children's vocabulary (% correct, B=.03, p=.014, ES=.14). Note, in these models we included the pretest score as a covariate to improve the precision of the estimated treatment effects. We will present final models for all outcome measures for both cohort 1 and 2 at the conference that include a full set of covariates, and that reflect the nested nature of the data (children in schools).

Conclusions:
These results provide preliminary evidence of the impact of the Word Generation intervention on middle school children using a measure designed to be proximal to the intervention, and with a small sample (cohort 1, including 6 schools). Additional analyses will examine treatment impacts on a larger battery of student assessments related to reading comprehension, including a substantially larger sample size and subgroup that was exposed to the intervention for two years. Implications will be discussed, including contributions to the evidence base for what works to promote greater proficiency in reading comprehension for high poverty schools.