

**Empirical Evidence on *The First 4,000 Words* Program: A Report
for Teachers Requesting Information on Research on the Program**

The First 4,000 Words program was developed with U.S. Office of Education IES/SBIR grant. This type of grant funds both development of educational materials and testing them. We conducted three studies on the program: a feasibility study conducted in Spring 2008, a pilot study conducted in Summer 2008, and a field trial conducted in Spring 2009.

Feasibility Study. The feasibility study (Sales, 2007; Sales & Graves, 2009) was conducted in two phases and yielded two sorts of information. In the first phase, first through fourth graders in three schools were given the 40-item *Seward Listening Vocabulary Assessment* and the 40-item *Seward Reading Vocabulary Assessment*. School 1 was a charter school in an urban setting with a large percentage of Hmong students. School 2 was the single elementary school in a small urban district abutting a large urban district and enrolled a large percentage of students of color. School 3 was the poorest performing elementary school in a large school district in a second ring suburb. A total of approximately 550 students took the test, including approximately 370 native English speakers, 130 English-language learners, and 50 special needs students. Results indicated that all but a few students could complete either version of the test in the 30 minutes allotted and that the tests met satisfactory levels of reliability and validity. While many children knew a lot of these very frequent words, some children knew few of them. For example, on the listening test, the bottom 10% of the first graders knew fewer than 4 of the 10 words representing the 1,000 most frequent words (see Table 1 for details).

Part two of the feasibility study was conducted at Schools 1 and 2, the schools with the largest numbers of English-language learners. Sixty-seven percent of the students in School 1,

and 26% of the students in School 2 were classified as English-language learners. Eight students from each school were recruited for this part of the study. Half of the students were labeled as LEP, they scored between the 25th and 75th percentiles of their schools on the Seward Vocabulary Assessments, and their teachers agreed that they would profit from additional vocabulary instruction. Observations indicated that although some changes to this initial version of the software were needed, it generally worked quite well and was well received by students. For example, most students found the program easy to operate, believed that they were learning the words taught, and were positive about doing more lessons.

Pilot Study. The pilot study (Fehr, Davison, Graves, Sales, & Seipel, 2009) was a brief intervention conducted in School 2. In this study, students studied 100 of the first 4,000 words using *The First 4,000 Words* program. The poorest-performing students on the Seward Vocabulary Assessments were randomly assigned to either a treatment group or a control group. Both groups took a 40-item pretest on the 100 words to be studied by the treatment group. Students in the treatment group (23 boys and 20 girls) left their classrooms and received the online vocabulary instruction in the library or resource room. They were initially introduced to the computer login and vocabulary lesson procedures and were provided ongoing help with these as needed throughout the study. Students typically spent 15-25 minutes per session and were able to complete 1-3 lessons during a session. Several experimenters monitored student progress, providing technical help and encouragement. Most students were able to complete the 10 lessons (100 words) comprising the treatment within 5-6 sessions (approximately 2 weeks). Students assigned to the control condition remained in their normal classrooms and received the normal curriculum. The week following the intervention, students took a post-test identical to the pre-test. In order to assess program feasibility, teachers were interviewed, and they completed a short

questionnaire after the completion of the study. Both the interviews and questionnaire (which used a 0 = poor, 1 = fair, 2 = good, 3 = excellent scale) asked teachers about their general impressions of the instruction and how well it would fit with the existing curricula.

Table 2 shows mean pre- and post-test raw scores and standard deviations for students assigned to the experimental and control conditions by grade level. Gain scores for the experimental groups were significantly higher ($p < .05$) than those for the control groups in grades 2 and 3, and overall. Teachers expressed generally favorable views, and felt that the program would integrate well with existing curricula and classroom frameworks. Teachers' ratings of overall perception of the program averaged 2.9 on the 3 point scale. Their ratings of how the program would fit within existing curricula averaged 2.8 out of 3.

Field Trial. In the field trial (Sales, Graves, Davison, Fehr, & Pakhomov, 2009), the curriculum was scaled up to include all $\approx 3,600$ words. First to fourth graders from three schools participated in the intervention for approximately three months. Students were randomly assigned to a treatment group or to a business-as-usual control group. Although the post-test means on the Seward Vocabulary Assessment for the treatment group were generally higher than those for the control group, few differences were statistically significant. This we believe was because the test assessed students' knowledge of all 3600 words in the program but the students had only studied a few hundred or fewer of them during the three months or less they were in the program. To get a more sensitive measure of growth, we have examined word and lesson level data. Analysis of over 2000 10-item lesson pre-tests and post-tests showed a mean pre-test score of 7.0 and a mean post-test score of 9.6, indicating that students learned the vast majority of words they were taught and achieved near perfect scores on the lesson post-tests. Moreover, these gains were reflected in a paper and pencil post-test on the words taught given at the end of

the study. A Chi square analysis showed that students who learned a word during a lesson (as indicated by an incorrect lesson pre-test response and a correct lesson post-test response) were twice as likely to get that word correct on the paper and pencil post-test as students who did not learn a word during a lesson (.66 vs .33), a difference significant at $p < .001$. Thus, students learned the words during the individual lessons and retained that knowledge on the post-test given weeks or months after the instruction.

For purposes of evaluating student engagement, we conducted extensive observations of students using the program. Engagement was demonstrated by direct attention to the computer screen and active participation in instructional activities. Over the course of the study, we collected data from approximately 700 time points. Students were judged to be appropriately engaged 91% of the time. Inter-rater reliability was .94. Importantly, approximately 300 time point observations conducted over the final month of the study showed students engaged 90% of the time, indicating that students remained highly engaged even after considerable use of the *First 4,000 Word* program.

Table 1. Mean Seward Listening Vocabulary Assessment Scores for the Bottom 5–10% of Students

Grade	N 5%/10%	Mean 5%/10%
1	7/13	12.57/14.15
2	7/14	14.57/17.64
3	6/12	11.17/21.50
4	7/15	23.29/25.07

Table 2. Mean Pre-test and Post-test Scores for Pilot Study 1

Group	Test	Control			Experimental			Effect Size
		n	Mean	SD ^a	n	Mean	SD ^a	
Grade								
2	Pre-test	9	10.6	3.2	8	10.9	5.5	1.7
	Post-test		10.1	3.5		18.3	9.1	
Grade								
3	Pre-test	6	21.7	7.3	8	23.3	6.7	1.3
	Post-test		24.3	8.6		30.5	5.6	
Grade								
4	Pre-test	6	19.5	2.5	6	18.3	4.1	0.9
	Post-test		24.3	9.5		27.3	9.0	
Total	Pre-test	21	16.3	6.7	22	17.4	7.5	1.1
	Post-test		18.2	9.9		25.2	9.4	

References

- Fehr, C., Davison, M. L., Graves, M. F., Sales, G. C., & Seipel, B. (2009). *The effect of individualized, online vocabulary instruction on picture vocabulary scores: A pilot study*. Poster session presented at the Fourth Annual IES Research Conference, Washington, D.C., June 8, 2009.
- Sales, G. C. (2007). *The First 4,000 Words: Phase I final report*. Minneapolis: Seward Incorporated.
- Sales, G. C., & Graves, M. F. (2009). Web-based pedagogy for fostering literacy by teaching basic vocabulary. *Information Technology, Education and Society*, 9 (2). 5-30.
- Sales, G. C., Graves, M. F., Davison, M. L., Pakhomov, S., & Fehr, C. N. (2009). *The First 4,000 Words: An interactive vocabulary development system for students in grades 1-4*. Poster session presented at the Fourth Annual IES Research Conference, Washington, D.C., June 8, 2009.