



BEHAVIORAL INSIGHTS  
AND PARENTING LAB  
THE UNIVERSITY OF CHICAGO

# **An Evaluation of the Big Word Club Vocabulary Program**

**FINAL REPORT**

**November 2018**

# **An Evaluation of the Big Word Club Vocabulary Program**

## **FINAL REPORT**

### **AUTHORS**

**Susan E. Mayer**, University of Chicago

**Ariel Kalil**, University of Chicago

**Philip Oreopoulos**, University of Toronto

The authors gratefully acknowledge financial support from J-PAL North America.

## **Acknowledgements**

Michelle Park Michelini was the lead Project Director for this evaluation and is responsible for its timely completion.

We are grateful to Shane DeRolf for his efforts not only in developing the Big Word Club but also in enthusiastically embracing the evaluation process. Rohen Shah and Harkirat Kaur contributed to the data analysis.

We are also grateful for the following members of the BIP Lab for important contributions to this evaluation: William Delgado, Nadav Klein, Keri Lintz, and Paula Rusca.

# TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY .....</b>	<b>6</b>
<b>I. INTRODUCTION.....</b>	<b>8</b>
<b>II. THE EVALUATION PROCESS.....</b>	<b>9</b>
II.A. SAMPLE SELECTION AND RANDOMIZATION.....	10
II.B. ATTRITION.....	10
<b>III. DATA COLLECTION AND MEASURES.....</b>	<b>11</b>
III.A. ASSESSMENTS.....	11
III.B. CHILD CHARACTERISTICS.....	14
III.C. SCHOOL CHARACTERISTICS.....	14
III.D. TEACHER’S USE OF THE BWC.....	15
III.E. TEACHER SURVEY.....	15
<b>IV. ESTIMATES OF PROGRAM EFFECTS .....</b>	<b>15</b>
IV.A. INTENT-TO-TREAT ESTIMATES.....	16
IV.B. TREATMENT-ON-TREATED ESTIMATES.....	17
IV.C. THE PPVT.....	17
IV.D. HETEROGENEITY EFFECTS.....	17
<b>V. THE ROLE OF THE BWC CONTENT ON PROGRAM EFFECTIVENESS .....</b>	<b>18</b>
<b>VI. THE ROLE OF TEACHERS’ USE OF THE BWC ON PROGRAM EFFECTIVENESS.....</b>	<b>19</b>
<b>VII. TEACHERS’ EVALUATION OF THE BWC.....</b>	<b>20</b>
<b>VIII. COMPARISON OF THE EFFECT OF BWC TO OTHER VOCABULARY PROGRAMS.....</b>	<b>21</b>
<b>IX. SUMMARY AND CONCLUSIONS .....</b>	<b>23</b>
<b>TABLES &amp; FIGURES .....</b>	<b>24</b>
TABLE 1: CHARACTERISTICS OF SCHOOLS BY TREATMENT (T) AND CONTROL (C) ASSIGNMENT AND SAMPLE.....	24
TABLE 2: CHARACTERISTICS OF STUDENTS IN TREATMENT (T) AND CONTROL (C) STUDENTS WHO TOOK THE 17- WEEK BWC ASSESSMENT .....	25
TABLE 3: CHARACTERISTICS OF STUDENTS IN TREATMENT (T) AND CONTROL (C) SCHOOLS WHO TOOK THE 25- WEEK BWC ASSESSMENT .....	26
TABLE 4: INTENT-TO-TREAT ESTIMATES OF THE EFFECT OF ACCESS TO THE BWC ON VALID BWC ASSESSMENT SCORES .....	27
TABLE 5: ITT-C ESTIMATES OF THE EFFECT OF ACCESS TO THE BWC ON VALID BWC ASSESSMENT SCORES FOR VARIOUS SAMPLES.....	28
TABLE 6: TREATMENT-ON-TREATED ESTIMATES OF THE EFFECT OF ACCESS TO THE BWC ON VALID BWC ASSESSMENT SCORES .....	29
TABLE 7: INTENT-TO-TREAT (ITT) AND TREATMENT-ON-TREATED (TOT) EFFECT OF ACCESS TO THE BWC ON AGE- ADJUSTED PPVT SCORES AT WEEK 25 FOR STUDENTS WITH VALID SCORES.....	30
TABLE 8: EFFECT OF ACCESS TO THE BWC ON BWC ASSESSMENT SCORES FOR VARIOUS SUBGROUPS.....	31
TABLE 9: TREATMENT TEACHER LOGINS TO THE BWC WEBSITE BY DATE .....	32
TABLE 10: TREATMENT TEACHER REPORTS OF THEIR USE OF THE COMPONENTS OF THE BWC PROGRAM .....	33
TABLE 11: REASONS TREATMENT TEACHERS DID NOT USE THE BWC.....	34
TABLE 12: TREATMENT TEACHER VIEWS ON THE CONTENT OF THE BWC.....	35
FIGURE 1: PERCENT OF CONTROL GROUP STUDENTS CORRECTLY IDENTIFYING EACH WORD ON THE BWC ASSESSMENT AT WEEK 17 .....	36
<b>APPENDICES .....</b>	<b>37</b>

APPENDIX A: SAMPLE EMAIL TO PRINCIPALS TO OBTAIN TEACHER PARTICIPATION APPROVAL.....	37
APPENDIX B: .....	39
<i>Figure A1: Minimum Detectable Effect Size for Varying Values of ICC</i> .....	39
<i>Figure A2: Minimum Detectable Effect Size for Varying Values of Treated Schools</i> .....	40
<i>Figure A3: Minimum Detectable Effect Size for Varying Number of Students</i> .....	40
<i>Table A1: ITT estimates of the effect of access to the BWC on the BWC Assessment for all students with valid scores controlling only state and student age for students with valid test scores</i> .....	41
<i>Table A2: TOT estimates of the effect of access to the BWC on the BWC Assessment controlling only state and student age for students with valid test scores</i> .....	42
APPENDIX C: THE BWC ASSESSMENT .....	43
<b>REFERENCES</b> .....	<b>45</b>

## Executive Summary

We evaluated the effectiveness of the Big Word Club (BWC), a web-based program of activities intended to help elementary school-aged children learn new vocabulary words by introducing one new word per day throughout the school year. We estimate whether pre-kindergarten and kindergarten children in schools randomly assigned to participate in the Big Word Club scored higher than students in a control group of schools on an assessment of receptive vocabulary based on words included in the BWC program (the BWC Assessment). We also assessed students using a standardized test of receptive vocabulary.

Teachers in classrooms with 862 students in pre-kindergarten and kindergarten in 53 schools volunteered to participate in the evaluation. The 53 schools were randomly assigned to either a treatment group in which teachers were given access to the BWC website or to a control group in which teachers were promised access to the BWC during the next academic year. After randomization six schools dropped out, leaving 818 students in 47 schools. Schools and students were balanced across treatment and control groups both before and after the six schools dropped out.

A time stamp provided information on each time a teacher logged on to the BWC website. Teachers had to log in to use the BWC lessons, although they could in principle log in for other reasons as well. Using the login data we estimate that about half of teachers in the treatment group logged into the BWC on at least half the available school days.

To take into account attrition, we used the average score of control group students and then the average score of the treatment group students to impute the scores of students in schools that dropped out after randomization. This provides realistic upper and lower bound on intent-to-treat estimates of the effect of the BWC.

The mean valid score on the BWC Assessment at 17 weeks was 24.1 words out of a possible 38 words. The minimum was seven words and the maximum was 37 words. Results show that children in classrooms with access to the BWC for 17 weeks identified, on average, between 1.194 and 1.319 more words on the BWC Assessment compared to the control group. The intent-to-treat effect size is between .229 and .267 standard deviations depending on how scores were imputed. All estimates are statistically significant at  $p < 0.01$ .

We used an instrumental variables model to estimate the effect of the BWC on students whose teacher logged onto the BWC at least once during the 17-week period (the treatment-on-treated model). In this model treated students identified 1.638 more words on the BWC Assessment at 17-weeks than the students who were not treated. The effect size was .320 standard deviations. Teachers report liking the BWC, believing that it is effective, and wanting to use it in the future. The main reason that they report not using the BWC is having too many other required activities.

We assessed students 25 weeks after the intervention began on words included in the first 17 weeks of the BWC program. At that point students in schools with access to the BWC identified

between 1.071 and 1.242 more words (depending on the imputation) on the BWC Assessment than students in the control schools. This corresponds to an effect size of between .229 and .257 standard deviations with all estimates statistically significant at  $p < 0.01$ . At 25 weeks the TOT estimated effect size was .305. This means that children in the treatment group retained their advantage over children in the control group for at least 8 weeks.

We also assessed students at 25 weeks after treatment using the Peabody Picture Vocabulary Test 4, a standardized test of receptive vocabulary. Children in classrooms with access to the BWC scored higher on the PPVT than students in classrooms without access, but the difference was not statistically significant at  $p = .05$ .

The estimated effect of the BWC was greater for female students compared to male students who took the 25-week BWC Assessment, but there was no difference in the effect by students' age, whether the student was in kindergarten, whether the student had English as a second language or had special needs, or whether the school was private or received Title 1 funding.

The effect of the BWC on receptive vocabulary is the same as or greater than the effect of the only two programs that we could find that are comparable to the BWC. The BWC may be more cost effective than other programs because it does not require teacher training or follow up.

# **An Evaluation of the Big Word Club Vocabulary Program**

Susan E. Mayer and Ariel Kalil, University of Chicago

Philip Oreopoulos, University of Toronto

## **I. Introduction**

Researchers, practitioners, and policymakers agree that vocabulary is a critical factor in literacy development, especially during early childhood (National Early Literacy Panel, 2008). The National Reading Panel (2000) identified vocabulary as one of five key aspects of literacy crucial to reading comprehension. Vocabulary is central to oral language development, reading comprehension, and development of domain-specific knowledge (Beck & McKeown, 2007; Cunningham & Stanovich, 1997; Scarborough, 2001; Snow et al., 1998; Stahl & Nagy, 2006; Roskos et al., 2008). Many studies show that the size of children's vocabulary is strongly related to how well they understand what they read (Stahl & Nagy, 2006; Spira, Bracken, & Fischel, 2005; Storch & Whitehurst, 2002). Early vocabulary powerfully predicts children's later language development, reading skills, school-readiness, and academic success (Dickenson & Porche, 2011; Morgan, Farkas, Hillemeier, Hammer, & Maczuga, 2015; Rowe, Raudenbush, & Goldin-Meadow, 2012). Research suggests that there are large differences in vocabulary knowledge between children from affluent and low-income backgrounds by the age of three (Farkas & Beron, 2004). By the end of Grade 2, on average children have acquired around 6,000 word meanings and those in the highest quartile have acquired around 8,000 word meanings. However, children in the lowest quartile have acquired only around 4,000 root words. This gap of 2,000 words is roughly equivalent to two grade levels in vocabulary development (Biemiller, 2005).

This evidence has led early childhood educators to strongly advocate for improvements in vocabulary learning in early childhood education. Much of current reading instruction is based on the premise that children build vocabulary after learning to read. Consequently, in many preschool and early elementary classrooms little or no vocabulary instruction occurs (National Reading Panel, 2000).

The Big Word Club (BWC) is a digital learning program that uses books, songs, animation, and dance to introduce children to a new word every day of the school year. To use the BWC, teachers log in to the BWC website where they access the lessons. For every week, the BWC website provides five videos arranged around a common theme that introduce each word for each day. It also provides one animated digital book, one animated music video, and one dance video—all of which include the five words for that week. The last video for each week includes a review of the week's words. Using the BWC requires no training for teachers. The website does not require teachers to use the lessons in the specified order and teachers can use some or all of the daily lessons for a week. Each video is 1–4 minutes long and the videos are made to be shared with the whole class at one time, so implementing the BWC is not costly in terms of classroom time. The BWC is intended to supplement, not to substitute, the normal classroom



literacy curriculum, and it provides flexibility to teachers to use the videos any time during the day.

The content is intended for children in preschool to Grade 5, with different classroom materials depending on the grade. In general, the words are “big” in the sense that many are not typical of the vocabulary of young children. For example, the words for preschoolers include “gargantuan,” “primate,” “prehensile,” “equator,” and “slither.”

The BWC is intended for classroom use over a school year or multiple school years. It provides activities for one word per day for 38 weeks of classroom use per year. It is available for schools or districts to purchase.

We evaluated the effectiveness of the BWC at increasing receptive vocabulary by comparing two measures of receptive vocabulary for students in schools with classrooms in which teachers volunteered to participate in the evaluation. Half of those teachers were randomly assigned to a treatment group that was given access to the BWC website. The other half of the teachers was assigned to a control group that was not given access to the website. Control group teachers were told that they would have free access to the BWC during the 2018–19 school year. All students were in either kindergarten or preschool.

In the next section we describe the evaluation procedures, including how we recruited and randomized schools and attrition from the sample. In Section III we describe the measures we used in the evaluation, including the assessments and school and child covariates. In Section IV we present the main results, comparing treatment and control students’ assessment scores. In Sections V and VI we describe two factors that are important to changes in vocabulary knowledge; namely, the words included in the BWC and the extent to which teachers used the BWC in their classrooms. In Section VII we describe teachers’ evaluations of the BWC. In Section VIII we compare the results of the BWC evaluation to the results of evaluations of other similar vocabulary programs and discuss the potential cost and benefit implications of that comparison. Finally, Section IX concludes the report.

## **II. The Evaluation Process**

Ideally, we would like to know whether students in schools that purchase the BWC experience greater growth in receptive vocabulary compared to students in schools that would have liked to purchase the BWC but for exogenous reasons, we were not able to. For practical reasons this is not what we tested in this evaluation. First, the schools were selected because a teacher in the school volunteered to participate in the evaluation at no cost to the school. Second, the evaluation tested the increase in receptive vocabulary over a 16-week period and not an entire school. Consequently, the mental experiment in this evaluation is whether, among a set of volunteer teachers, students in the classrooms of those who were given access to the BWC for 16 weeks scored higher in receptive vocabulary compared to students in the classrooms of those who were not given access to the BWC.

**II.a. Sample selection and randomization.** In order to recruit schools to participate in the evaluation, Shane DeRolf, founder of the BWC, launched a campaign on the Big Word Club's Facebook page that asked teachers to volunteer. Interested teachers were provided with a link to information about the evaluation and how to apply to participate. From this campaign, 637 teachers applied to participate.

Of these teachers, 260 were in schools that clustered in Arizona, Colorado, and Texas with the remaining schools scattered throughout the United States. Because we had to travel to schools to assess students, it was cost prohibitive to include schools that were very geographically dispersed. Thus, we concentrated on schools in these three clusters. From the 260 classrooms, we eliminated schools in which all of the students in the volunteering classroom had special needs because we did not have staff qualified to assess special needs students. We also eliminated schools that were more than 100 miles outside of the main cities where the schools clustered in each state; namely, Denver, Phoenix, Houston, and San Antonio. Finally, we eliminated home-based childcare centers with a very small number of children. This left 96 eligible schools in which at least one teacher had volunteered.

We next emailed the principals of these 96 schools to tell them about the evaluation and to ask if they agreed that teachers could participate. The email is in the appendix (see Appendix A). From these 96 emails, 53 principals agreed that their school could participate in the evaluation. The main reason principals gave for not participating was either that the school district prohibits schools from participating in outside research projects or that the school district requires a lengthy application process for outside researchers. Of the 53 schools that agreed to participate, ten were in Arizona, 20 were in Colorado, and 17 were in Texas. We randomized these 53 schools within states to either the treatment or the control group.

For the initial power estimates, we assumed 52 schools with 16 students each (for a total of 832 students), with half the schools assigned to treatment. We also assumed an intraclass correlation of 0.2 and set power to 80% and significance level to 5%. Given these parameters the minimum detectable effect size was 0.202. See Appendix B for more information on the power estimates.

Schools in the treatment group were given access to the BWC starting on November 13, 2017. This was the earliest date in the school year that we could begin the intervention given the length of time it took to recruit schools and to provide information to principals about the assessment process. We first assessed students around 17 weeks after treatment teachers were first given access to the BWC. In order to test for short-term retention of BWC words, we implemented a second assessment that began 25 weeks after treatment group teachers first had access to the BWC. This assessment included only the words that were on the 17-week assessment even though teachers continued to have access to the BWC website after the 17-week assessment. The 25-week assessment took place in May, 2018, just as some schools were about to begin summer vacation. This was the latest we could do the assessment within the same academic year as the first assessment.

**II.b. Attrition.** After randomization, six schools dropped out. Three of these schools were in Colorado and three were in Texas. One of these schools was in the treatment group and five were in the control group. This left 47 schools in the sample. The reasons that schools gave for

dropping out after randomization were scheduling conflicts with the assessment period, long-term teacher absence due to health issues, unwillingness to provide the research team with student names and birthdates, and unwillingness to allow external personnel to assess students.

Table 1 shows characteristics by treatment status of the 53 schools that were originally randomized and the 47 schools that remained after six dropped out. All six schools that dropped out are public schools, and half are Title I schools. Compared to the 26 schools that were randomized to the control group, the remaining 21 schools in the control group are more likely to be private schools and less likely to be Title 1 schools. However, of the three characteristics for which we have data on all schools, there is no statistically significant (at  $p=.05$ ) difference between treatment and control schools.

### III. Data Collection and Measures

In this section we describe the data used in the evaluation. This includes two measures of receptive vocabulary, school and student characteristics, and data from a survey of both control and treatment group teachers. In addition, the BWC provided information on teacher logins to the BWC website, which we use as a proxy for teacher use of the BWC.

**III.a. Assessments.** The first goal of the BWC is to increase vocabulary by familiarizing children with the specific words included in the BWC. Consequently, we assessed the extent to which children in classrooms with access to the BWC were able to identify words included in the BWC program compared to children in classrooms with no access to the BWC. A second goal of the BWC is to increase children's curiosity about and receptivity to words not specifically included in the BWC and hence to increase their general vocabulary. Consequently, we also assessed children's overall receptive vocabulary using a standardized and widely used assessment of receptive vocabulary: the PPVT Form A (henceforth referred to as the PPVT).

We developed the BWC Assessment to be similar in presentation to the PPVT. For each word, a child was shown a paper with four pictures. One of the pictures depicted the target word and the other three pictures depicted something else. The child was asked to point to the picture that depicted the target word. The child received a point for each correctly selected word, and the score on the assessment was the total number of correctly selected words. The complete BWC assessment and score sheet are included in the appendix (see Appendix C).

We selected words from the first 16 weeks of the BWC list to include in the BWC Assessment. Although in principle teachers had access to the BWC for 17 weeks at the time of the first assessment, due to holidays and other days off school, most teachers had the opportunity to use fewer than 17 weeks of the program. From the list of the 80 words in the first 16 weeks of the BWC, we had to omit several because it was not possible to depict them with a picture. For example, the adjective "silent" is difficult to depict in a static image. Some words also had multiple possible meanings, making them difficult to depict. For example, the adjective "healthy," can be interpreted as "in good health" or "good for you." We also eliminated some of the names of dinosaurs included in the BWC because it was difficult to show pictures that clearly distinguished one kind of dinosaur from another. After eliminating 20 difficult-to-depict words,

we randomly selected 38 words from the remaining 60 words to include in the BWC Assessment. We selected 38 words because a pilot of a 20-word version of the assessment indicated that children could do the assessment rapidly, and also because the attention span of the youngest children was short. Children were able to complete the 38-word assessment in an average of about five minutes. Children were tested on all 38 words whether or not their teacher had covered the words in class.

We administered the BWC Assessment after teachers had access to the BWC website for 17 weeks (we call this the 16-week assessment because we tested students on words included in the first 16 weeks of the BWC program). We administered the same BWC Assessment after teachers had access to the BWC for 25 weeks. We refer to that assessment as the 25-week assessment. It was intended to test retention of words learned during the first 16 weeks of the BWC. The BWC is intended to be used over an entire school year or even over multiple school years. However, in order to get both an initial assessment and a test of retention in the same school year, we had to assess learning over a shorter period of time. We first assessed students after teachers had access to the BWC website for only 17 weeks.

The PPVT is designed so that the easiest words are in the beginning; the words get progressively more difficult. Children start the assessment at the age-appropriate item set. There are 12 items in a set and 19 sets in total on the assessment, but even six-year-old children rarely get past set 12 or 13. A basal set is established when one or zero errors are made in a set. The test moves forward until the student makes eight or more errors in a set. At this point the assessment ends.

The BWC Assessment was administered as a test booklet with paper score sheets for assessors to complete. The PPVT was administered on iPads through Pearson's Q-Interactive digital platform. Q-Interactive provides automated score reports with raw and age-adjusted standardized scores.

The PPVT took an average of ten minutes for children to complete. Because the PPVT is adaptive to students' responses, the duration of the assessment is shorter when more responses are incorrect.

We administered the PPVT at the same time as the 25-week BWC Assessment. We administered the PPVT after students had taken the BWC Assessment. Changes in a general measure of receptive vocabulary—such as the PPVT—are expected to lag changes in the recognition of the words that are included in the BWC, which are expected to occur simultaneously with the lesson in which the word occurs. Changes in general vocabulary are expected to occur as a result of access to the BWC if students become more interested in learning new words when the experience of learning new words has been enjoyable. In addition, gains in vocabulary increase listening comprehension, which provides greater context for learning new words, further increasing vocabulary. As children learn new words they develop phonological representations of words, which allows them to decode new words more quickly (Ouellette, 2006; Wise et al., 2007; Gathercole, Service, Hitch, Adams, & Martin, 1999).

Assessors were not told whether they were assessing students in a treatment or a control school. Assessors were asked to interact with students prior to testing and to record any issues related to

language, or cognitive, or behavioral development. Assessors also noted how distracted the child was during both the BWC Assessment and the PPVT. Distraction levels were coded from 0–3, with 0 being no prompts needed to redirect the child’s attention to the assessment and 3 indicating that at least three prompts were needed to refocus the child. Assessors also noted test duration and any unusual circumstances that occurred during test administration.

Assessors were instructed not to show, spell, define, or use the stimulus word in a sentence or use “a,” “an,” or “the” before the word. They were instructed to use neutral language like “okay” and “thank you” rather than “good job” or “well done.” Comprehensive training on human subjects research and assessment administration was provided to assessors. The training protocol for BWC Assessment and PPVT assessment administration is available upon request.

We did not assess children in the six schools that dropped out after randomization. In the remaining 47 schools, we assessed an average of 17.4 students in each of the 47 schools that participated in the evaluation. The minimum number of students assessed at a school was 13 and the maximum was 26. We assessed a total of 818 students across all schools.

Thirty-six of the 47 schools had only one classroom participating in the evaluation. When a school had only one participating classroom, we attempted to assess all students in that classroom. Assessors selected students for assessments in the order in which they appeared on a roster that listed students in random order by an ID number. Because we could visit schools only once, a student was skipped if that student was absent or otherwise not available when it was his or her turn to be assessed. There were 797 students in schools with only one classroom and we assessed 638 of those students, including 348 in 19 treatment schools and 290 in 17 control schools.

Eleven schools had more than one participating classroom. In five of these schools, we assessed children from only one randomly selected classroom. In two schools, administrators provided to us a student roster that included more than one classroom. We provided the assessors with the list of students in random order and they selected students in the order that they appeared. Overall, we assessed children in all but four participating classrooms.

Our procedure for validating BWC Assessment scores was the same as the procedure described in the PPVT manual, with two exceptions. First, we invalidated scores for children who were not enrolled in the classroom by January 1, 2018, because these students would have missed a significant proportion of the potential time that teachers could use the BWC. Second, as we describe below, there were a few scores on the PPVT invalidated because of the errors in how the scores were entered into the electronic scoring platform. Otherwise, we classified scores as invalid if

- the assessor noted the child had serious difficulty understanding directions for completing the assessment;
- the assessor noted the child was uncooperative;
- the assessor noted the child received strong positive or negative reinforcement from school staff during the assessment;
- the assessment was incomplete.

For the 17-week BWC Assessment of 818 students, there were 46 invalid scores. Of these, 23 were for students in treatment schools and 23 were for students in control schools. The average invalid score on the BWC Assessment was 17.8 for students in treatment schools and 15.4 for students in control schools. This difference is not significant ( $p=.05$ ).

For the 25-week BWC Assessment, we only included students who had valid test scores in the first round because the intent was to assess retention. Therefore, our total eligible sample for the 25-week assessments was 772. Out of that number, we assessed 603 students with the BWC Assessment and 602 students with the PPVT. We visited each of the 47 schools on two separate days. We assessed an average of 13 students in each of the 47 schools. The minimum number of assessed students at a school was nine and the maximum was 17. The 25-week assessment occurred close to the end of the school year, which presented challenges related to scheduling the assessment, higher absenteeism, and lower focus among students (meaning that more children were unavailable to be assessed and more were uncooperative). Some children were not able to take both assessments; there were seven students who took the BWC 25-week assessment who did not take the PPVT and six students who took the PPVT who did not take the BWC Assessment.

For the 25-week sample of 603 students, there were six invalid scores with four in treatment schools and two in control schools. The average BWC Assessment score was 23.5 for invalid scores in the treatment schools and 19.0 for the invalid scores in the control schools.

As noted above, we administered to PPVT using a digital platform on an iPad. In seven cases a student response did not register on the iPad. PPVT scores were invalidated for the same reasons that we invalidated scores on the BWC. Altogether, 15 PPVT assessments were invalid. Six of these were in treatment schools. The difference in the invalid score for the treatment group (66.0) and the invalid scores for the control school (66.5) was not statistically significant at  $p=.05$ .

**III.b. Child characteristics.** At the time of the 17-week BWC Assessment we collected data on individual children including their names, gender, and date of birth. A few schools provided data on whether children were learning English as a second language or whether they had special needs. However, most schools did not provide this information. Because these factors are likely to be important influences on children's ability to learn new vocabulary words, we asked assessors to note whether children had observable developmental problems likely to indicate special needs and whether the child could not communicate fluently in English. Assessors marked children who were not fluent in English as "observed ESL" (OESL) and children who appeared to have serious developmental problems as "observed special needs" (OSN).

**III.c. School characteristics.** We collected data on several characteristics of the schools in our sample primarily from the United States Department of Education's National Center for Education Statistics (NCES) or from school district websites. These data include whether the school was public, private, or charter, whether it was designated as a Title 1 school, level of enrollment, and the percent of students in the school eligible for free or reduced lunch. When information was not available from published sources we asked school principals for the information.

**III.d. Teacher’s use of the BWC.** When a teacher logs in to the BWC website, a time stamp allows the program to generate data on the number of times a teacher’s login information was used to log in to the BWC. This usage data tracked the number of logins but not the time or duration of logins. Teachers could log in for multiple reasons besides using the BWC program in class. For example, if a teacher viewed the BWC materials outside of class time it will count as a login. Additionally, if a teacher allowed others to log in with his or her information, it would appear that the teacher had logged in. However, we use the number of logins as an approximation of teacher use of the BWC program.

**III.e. Teacher survey.** At the time of the 25-week assessments, we gave teachers from both the control and treatment groups a survey that asked about characteristics of students in their classroom. The treatment group teachers were additionally asked about their usage of the BWC program and their opinions about several aspects of the program. Of the 67 eligible teachers, we received survey responses from 66 teachers. All of these were the lead teacher in their classroom. This included teachers whose students we assessed and teachers whose students we did not assess. Although 66 teachers returned surveys, only 51 returned completed surveys.

## IV. Estimates of Program Effects

The mean valid score on the BWC Assessment at 17 weeks was 24.1 words out of a possible 38 words. The minimum was seven words and the maximum was 37 words.

Table 2 shows characteristics of students in the treatment and control groups for all students and for students with valid test scores who were assessed in week 17. Forty-six students had invalid test scores. Omitting them increases the percent of students in kindergarten slightly and reduces the percent of students classified as special needs slightly. However the reductions were similar for the treatment and control groups and none of the differences between the treatment and control group was statistically significant at  $p=.05$ .

Table 3 shows the same information for students assessed at 25 weeks. We only assessed students at week 25 who had valid scores at week 17. We were able to assess 603 of the 772 students (78.1%) with valid scores at week 17. Six additional students had invalid scores at 25 weeks. Students’ gender, grade, and OELS and OSN designation were the same at 17 weeks as at 25 weeks, so differences in the percent of students with these characteristics between the assessments is primarily due to selection in who took the 25-week assessment. None of the differences across students in treatment and control schools was statistically significant at  $p=.05$ .

We estimated a standard model of the average intent-to-treat (ITT) effect including a set of covariates:

$$(1) \quad S_i = \beta_0 + \beta_1 T_i + \beta_2 A_i + \beta_3 CO_i + \beta_4 TX_i + \beta_4 X_i + \beta_4 X_S + \varepsilon_i$$

where  $S$  is a score on one of the assessments,  $T$  is an indicator for random assignment to a school in the treatment group,  $\varepsilon$  is an idiosyncratic error term, and  $i$  represents the individual in the school.  $A$  is an indicator variable for the child’s age because age is perhaps the most important determinant of vocabulary and the BWC assessment is not age adjusted.  $CO$  and  $TX$

are state indicators for whether the school is located in Colorado or Texas respectively (making Arizona the default state in the regression), included because we randomized schools within states.  $X_i$  represents a set of baseline characteristics of students that are likely to affect the acquisition of vocabulary. These include whether the child is female, whether the child is in kindergarten, whether the child is observed to be an English language learner, whether the child has observed special needs.  $X_s$  represents a set of school characteristics likely to influence the acquisition of vocabulary. Table A1 shows model 1 without the students and school covariates. These include the proportion of the children in the school receiving free and reduced lunches, whether the school is private, and whether the school is a Title 1 school. In the ITT models, standard errors are corrected for clustering in 53 schools. Our parameter of interest is  $\beta_1$ , which equals the average difference for children randomized to the treated group compared to those randomized to the control group.

**IV.a. Intent-to-treat estimates.** As we have noted, six schools dropped out after randomization, one in the treatment group and five in the control group. These schools had 90 students. We do not have test scores for these students because we could not assess them. To estimate intent-to-treat effects we bound the possible range of these scores by first assuming that all 90 students would have had the average score of the students with valid scores in the control group had they remained in the sample. This assumes that students in neither the control schools nor the treatment schools that dropped out would have learned any BWC words during the intervention. We call this the ITT-C estimate to indicate that the scores of students in the schools that dropped out were imputed with the mean score of control group students.

Second, we assume that all the students in the schools that dropped out would have had the mean score of the students in the treatment group had they remained in the sample, assuming that students in the control group would have learned the same number of words from the BWC program as the students in classrooms with access to the BWC website. We call this the ITT-T estimate to indicate that these scores were imputed using the mean score of the treatment group.

Table 4 shows the results of the ITT-C and the ITT-T estimates of the effect of access to the BWC on the BWC Assessment for students with valid test scores. The row labeled “treatment” shows how many more words students in treatment schools identified compared to students in control schools. Students in the treatment group identified between 1.194 and 1.319 more words than students in the control group on the 17-week BWC Assessment. We computed effect sizes as the ratio of the treatment coefficient to the control group standard deviation. The additional words scored by the treatment groups corresponds to an effect size of between .256 and .284 standard deviations. All estimates are statistically significant at  $p < 0.01$ .

Table 4 also shows that on the 25-week BWC Assessment the treatment group correctly identified between 1.071 and 1.242 more words than the control group; effect sizes are between .229 and .267 standard deviations.

The 25-week BWC Assessment was the same as the 17-week BWC Assessment because we intended to assess the retention of words learned in the first 16 weeks of the BWC program. As noted above, not all students who participated in the 17-week BWC Assessment also participated in the 25-week BWC Assessment, so we cannot compare the 25-week scores to the 17-week



scores on the BWC Assessment because they are different samples. In Table 5 we reproduce the results for the ITT-C estimates at 17 and 25 weeks but we add the ITT-C estimate for only students who had scores for both assessments. It shows that for students who took both assessments, treatment students scored an additional 1.509 words on the 17-week assessment compared to the 1.319 additional words for the entire sample of students who took the 17-week assessment. This means that the students participating in the 25-week assessment are positively selected on vocabulary knowledge compared to the whole sample. However, in either case, the 25-week assessment scores are very close to the 17-week assessment scores,.

**IV.b. Treatment-on-Treated Estimates.** We also estimate the effect of participating in the program, or the effect of the treatment-on-treated (TOT), using random assignment to treatment as an instrument for actual participation (Bloom, 1984; Angrist, Imbens, & Rubin, 1996). We use two-stage least squares to estimate the TOT with the instrument being an indicator for schools in which a teacher implemented the BWC program (that is, the teacher logged on to the BWC website at least once), and 0 for the schools in which the teacher did not implement the BWC (that is, the control group and schools whose teacher never logged on to the BWC website).<sup>1</sup>

Table 6 shows that among students with valid scores, treated students whose teacher logged on to the BWC website correctly identified 1.638 more words on the BWC Assessment at the 17-week mark compared to the students in the control group. The effect size is .320 standard deviations. At 25 weeks the treated students with valid scores identified 1.561 more words than untreated students. The effect size is .305 standard deviations.

**IV.c. The PPVT.** Table 7 shows the ITT and TOT estimates for the PPVT given at 25 weeks for students with valid test scores. Overall, the mean age-adjusted PPVT score for treatment and control students was 99.5, which indicates that students scored at about the national average. The ITT-C estimate shows that students with access to the BWC scored .151 standard deviations higher on the PPVT compared to students without access to the BWC. The estimated 95 percent confidence interval of the true effect includes the point estimates for the ITT-C and ITT-T effects from the BWC Assessment, but also includes zero. Thus the PPVT estimates are not sufficiently powered to rule out important or null impacts. The same is true for the PPVT ITT-T and TOT estimates.

**IV.d. Heterogeneity effects.** To check for heterogeneity in treatment effects, we estimated a series of models like the ITT-C model described above, but in each we included an interaction term for being in the treatment group and one of several characteristics of students or schools.

---

<sup>1</sup> This estimate assumes that no control group teachers accessed the BWC program. Because control group teachers could not access the BWC website and the login data discussed below shows that no control group teacher did log in to the BWC website, this assumption seems warranted. The TOT model is as follows. First stage:

$$(1) U = a_0 + a_1T + a_2Age + a_3 CO + a_4 TX + u.$$

Second stage:

$$(2) Y = \beta_0 + \beta_1U + \beta_2 Age + \beta_3 CO + \beta_4 TX + \varepsilon.$$

The coefficient on the interaction terms estimates the difference in the treatment effect for students with that characteristic compared to students without that characteristic.

Table 8 shows the treatment coefficient and the coefficient on the interaction term. Only the interaction coefficient for being female was significant at  $p=.05$  and only for the 25-week BWC Assessment. That coefficient suggests that female students were more responsive than male students to exposure to the BWC.

## **V. The Role of the BWC Content on Program Effectiveness**

The extent to which children learn new words from a vocabulary program depends on several factors. We focus on two factors that could influence the success of the BWC. The first, which we discuss in this section, is the content of the program. The second, which we discuss in the next section, is how often the program is used.

For students to learn new words from an intervention, they must not know the words before the intervention begins. If, for example, all the words in the BWC were known by students before they began the BWC, we would observe no difference between students in the treatment and control schools at the end of the intervention. If students knew none of the words before being exposed to the BWC, we would expect students in the treatment but not control schools to learn words from the BWC (unless control group students were exposed to the words somewhere else). Thus, the fewer words known by students before the intervention begins, the greater the chance that the students in the treatment group will know more words than students in the control group at the end of the intervention.

The BWC includes words of varying familiarity for preschool and kindergarten children. We were not able to do a baseline test of knowledge of the words in the BWC, but a reasonable proxy for word familiarity in this age group is the number of words that students in the control group could identify at the end of 17 weeks, since we know that control group students were not exposed to the BWC and probably had limited opportunity to learn the specific words in the BWC from other sources. Figure 1 shows that over half of the students in the control group were able to correctly identify 22 of the 38 words (58%) in the BWC Assessment and over 75% of control group students were able to identify nearly a third of the words.

The correlation between the percent of control group students who know a word and the difference between treatment and control students who know the word is  $-.43$ , which indicates that the more a word was known to control group students, the smaller the difference between the percent of treatment and control students who correctly identify the word. For eight of the ten words most frequently known by control group students, there was no statistically significant ( $p=.05$ ) difference between treatment and control group students knowing the word at the 17-week assessment. On the other hand, treatment students were significantly ( $p=.05$ ) more likely to know eight of the ten words least likely to be known by control students.

The difference between the number of words that the treatment and control group can identify after an intervention depends importantly on the number of words that were known to both

groups before the intervention. The effect of the BWC on the BWC Assessment score may have been greater had it included more words that were unknown by all students at the start of the intervention. As we discuss below, because most vocabulary programs are evaluated by testing the words that are included in the program, keeping in mind that the choice of words is crucial to the effect of a program is useful for comparing across interventions.

## **VI. The Role of Teachers' Use of the BWC on Program Effectiveness**

Even a very good program cannot help children build vocabulary if teachers do not use it. A teacher must log in to the BWC website and use the BWC. As described above, we have a count of each time a teacher did so. For the reasons described above, the usage data does not perfectly capture a teacher's use of the BWC in the classroom. However, it is still informative. First, there is no record of any teacher in the control group logging on to the BWC website, which suggests that there was no contamination of the control group. Second, the login data provides a rough estimate of how much teachers used the BWC.

We received login information at four different times during the evaluation period. Table 9 shows how many logins were recorded for each time period. By the end of week 25 (when we received the last login information), teachers had logged on an average of 61.4 times. The maximum number of logins was 196 and the minimum number of logins was zero, indicating a great deal of variation in usage.

Teachers were more likely to log in at the beginning of the intervention period. The maximum potential school days is the number of school days there would have been, had school been in session all five days of every week for each period. Because of holidays, teacher training days, and other days off, most teachers were not in the classroom for all of these days and the actual number of school days varies by school. The actual number of days probably averages about 90 percent of the potential days. Up to May 7, which was when we began the second assessment, there had been 112 potential school days since the start of the intervention. The last row in Table 9 shows the mean number of logins per potential school days. If all teachers used the BWC for every potential school day, this number would equal one. Because the number of potential school days is greater than the number of actual school days, we expect this number to be less than one. In the first period for which we received login data, teachers on average logged in on 73 percent of potential school days. In each period the number declined, so that in the last period they logged in only 54 percent of potential school days. There are many potential explanations for the decline in logins. For example, it is possible that in the beginning teachers logged in when they were not in the classroom in order to explore and learn about the program. The last period was very near the end of the school year, so there were likely to be more activities that competed for classroom time. Nonetheless, the login data suggest that teachers used the BWC half or fewer of the days that it was intended to be used.

Out of 44 teachers with recorded logins, by May 7—when there had been 112 potential school days—20 teachers had at least 50 logins, 16 had at least 75 logins, and 11 had at least 100 logins.

We cannot tell how many more of the BWC words children would have learned or how much better they would have done on the PPVT with greater usage.

The BWC has five components: a video on the word of the day, a book, a video with a song about the word for that day, a video with a dance for the word of the day, and a review of the words for a week. We asked teachers how much they used each component. Their responses are summarized in Table 10. Consistent with the usage data, most teachers reported using each component one or two times per week. However, a little over a third of teachers reported using the Word of the Day every day; judging from the login data, this is probably an overestimate of how often teachers used the BWC. Teachers reported using all the other components much less often.

Overall, the usage data suggests that compliance with the intended use of the BWC was moderate but with high variation in the intensity of usage.

## **VII. Teachers' Evaluation of the BWC**

Teachers may fail to use an educational program for a variety of reasons. For example, if teachers have to log in to a website to use a program but the classroom has inadequate bandwidth or other technical problems, their usage will be reduced. If teachers do not think students will learn from the program or think that the program is otherwise not useful, teachers will also be less likely to use it. If teachers have a lot of competing responsibilities, that will also likely result in diminished usage.

We asked treatment group teachers to report why they did not use the BWC on the days that they did not use it. Table 11 shows the response from the 35 treatment group teachers (out of 43 treatment teachers who were eligible to respond) who answered this question. The majority said that the main reason for not using the BWC was that they had too many other required activities. The second largest category was "I forgot," which suggests that one way to increase teacher use of the BWC is simply to remind them to use it.

We also asked treatment group teachers how many words they thought that their students had learned from the BWC. Over half of teachers thought their students had learned between six and 15 words, and another 12 percent thought their students had learned more than 15 words. Clearly, teachers were optimistic about the efficacy of the BWC, since they greatly overestimated the number of words that students learned. Consequently, it is not surprising that when asked how interested they were in using the BWC in the future, about 75 percent said that they would probably or definitely be interested.

We asked treatment group teachers several questions about the content of the BWC. These results are shown in Table 12. Few teachers thought that the vocabulary words were too easy, but close to half thought that the words were a little or much too difficult. This is surprising given the large proportion of control group students who appeared to know many of the words. The majority of the treatment group teachers thought that the length of the BWC was about right, and all but 6 percent thought that was appealing.

Although compliance with the intended use of the BWC was only modest, this does not appear to be because teachers were not interested in using BWC or that they did not think it would be useful. Instead, it appears to be because of competing classroom activities and sometimes that teachers forgot to use it.

## **VIII. Comparison of the Effect of BWC to Other Vocabulary Programs**

Many studies have estimated the effectiveness of vocabulary interventions on vocabulary acquisition and associated literacy skills such as reading comprehension, including several meta-analyses (for a summary of meta-analyses of vocabulary programs, see Marulis & Neuman, 2010). Two recent meta-analyses (Marulis & Neuman, 2010; Elleman, Lindo, Morphy, & Compton, 2009) cover a large number of studies using either experimental or quasi-experimental techniques. Unfortunately, this rich previous literature does not provide very useful benchmarks for the BWC effects. In their meta-analysis, Marulis and Neuman include interventions with children up to high school, and vocabulary acquisition among older children is likely to differ significantly from vocabulary acquisition among younger children. Both meta-analyses include a large number of small studies implemented by researchers to answer theoretical or practical questions about how to teach vocabulary. These programs are not intended for classroom use. Many of the studies omit from their analyses schools, classrooms, or children who did not fully comply with the implementation protocol, so the results are not useful for understanding what would happen in real classroom settings. Many of the studies included in the meta-analyses have flaws in the statistical techniques that diminish their usefulness.<sup>2</sup> Finally, as we have noted, the effect size of an evaluation depends crucially on the type of words that children are expected to learn, which makes it difficult to know when programs are comparable.

We were able to find only two vocabulary programs that provide a good comparison to the BWC in that they are intended for classroom use, are intended to supplement rather than replace the existing literacy curriculum, are directed at young children, and are evaluated using randomized controlled trials.

The Elements of Reading program was evaluated in 44 moderate- to high-poverty elementary schools over two consecutive school years (Apthorp et al., 2012). Children in kindergarten, first, third, and fourth grade participated. Control teachers were asked to provide literacy instruction in a business-as-usual manner. Treatment teachers were asked to implement the Elements of Reading program, which includes structured lesson plans and oral and written language activities for six to eight words per week, providing multiple exposures and opportunities for using the words over the 24 weeks of the intervention. It focuses on Tier 2 words. School district reading or language arts coordinators served as study coordinators, and at each participating school, assistant principals coordinated data collection. Researchers trained reading coaches in each school. The reading coaches helped train and support treatment teachers. Researchers observed

---

<sup>2</sup> Some of the flaws in statistical techniques include not correcting standard errors for clustering, and randomizing on the outcome.

teachers in the classroom to assess compliance with the Elements of Reading curriculum.

The high level of district and school engagement led to very high compliance with the intended implementation of Elements of Reading. Teachers self-reported using the program for the intended number of lessons on average, and classroom observations showed that on average, teachers implemented 85–91 percent of the intended activities depending on the grade. At the end of the intervention students were given an assessment of words included in the Elements of Reading curriculum. After the first year of the two-year intervention, the effect size for vocabulary knowledge on this assessment was .98 for kindergartners and 1.00 for first graders. However, after children had participated in the program for two years the effect on this vocabulary assessment was very small and not statistically significant. Nor was the effect on the Group Reading and Diagnostic Evaluation (GRADE)—a standardized assessment—statistically significant.<sup>3</sup>

Another study tests the impact of PAVEd for Success (K-PAVE) on kindergartners' expressive vocabulary in districts and schools in the Mississippi Delta (Goodson, Wolf, Bell, Turner, & Finney, 2010). The K-PAVE program includes instruction on a large set of thematically related target words through the provision of definitions, examples, and visual images of the words; through embedded instruction using storybook reading, extension activities, and teacher conversation; through interactive book reading to build vocabulary and comprehension skills; and through directed adult-child conversations to build vocabulary and oral language skills. Sixty-five volunteer schools participated in the evaluation. Teachers in the treatment condition received two days of initial group training in fall 2008, three follow-up telephone conference calls to discuss implementation issues and reinforce key aspects of the K-PAVE program, and two rounds of classroom observation and feedback on how to improve their implementation of K-PAVE. Fidelity of K-PAVE implementation was evaluated using a rating system provided by the program developer and administered based on classroom observation. Researchers report that there was substantial variation in fidelity of implementation across classrooms, but that it was implemented with “sufficient fidelity to support impacts on students.” The estimated impact of K-PAVE on expressive vocabulary as measured on the Expressive Vocabulary Test 2<sup>nd</sup> Edition—a standardized test—was an effect size of 0.14, which was statistically significant at  $p=.05$ .

The effect size for the students assessed with the BWC Assessment is greater than the effect size for either the K-PAVE program or the Elements of Reading program. The K-PAVE evaluation assessed students using a standardized test of expressive vocabulary, and in virtually all research on vocabulary interventions effect sizes are smaller for standardized tests than for tests developed by researchers to assess knowledge of the words included in the program. The effect size for the PPVT in the BWC evaluation was of the same magnitude as the effect size for the K-PAVE intervention, but the K-PAVE sample was larger than the BWC sample.<sup>4</sup> Elements of

---

<sup>3</sup> The published paper is not clear on the extent of attrition, reasons for attrition, or how attrition was handled analytically. Thus, it is unclear whether these estimates are ITT or TOT estimates.

<sup>4</sup> To put the PPVT effect size in context, many well-known interventions that provide a literacy curriculum have found no statistically significant effect of the intervention on PPVT scores. These include Doors to Discovery (Christie, Roskos, Vukelich, & Han, 2003), Even Start, Bright Beginnings (PCER Consortium, 2008), and Ladders to Literacy (PCER Consortium, 2008). Consequently, changing PPVT scores is a high bar to set for any intervention.

Reading was implemented with a very high degree of fidelity and K-PAVE was implemented with fidelity that was probably more similar to the fidelity of implementation of the BWC.

It is impossible to do even a back-of-the-envelope cost-benefit comparison of the BWC and other programs. The current cost of the Big Word Club to schools is about \$60.00 per classroom, regardless of the size of the class. The Big Word Club's stated mission is to help bridge the word gap. To that end, for each Big Word Club subscription sold, the company's "From One to Many" program will give a subscription to classrooms in disadvantaged schools. The costs of both the K-PAVE and Elements of Reading interventions are likely to be greater than the cost of the BWC because both require extensive teacher training and follow-up and because the classroom observations to assess fidelity may themselves be an important program component to get teachers to use the program. The BWC is likely less expensive and similarly effective, even with modest fidelity compared to these other programs.

However, even if we had a good estimate of the costs of the BWC, we would not be able to estimate the benefits of the program because we do not know the benefit of learning the words in the BWC. Comparing the BWC to K-PAVE and Elements of Reading shows that they both required substantial teacher training and ongoing support during the intervention, which is likely to raise their costs.

## **IX. Summary and Conclusions**

Over a 17-week period, access to the BWC web-based program increased recognition of the words included in the BWC program by over 0.250 SD. This increase is similar to or larger than the effect size for either of the two vocabulary interventions that we found that are similar in intent to the BWC. The estimated effect of the BWC on the PPVT 0.150 SD with a confidence interval that includes both zero and 0.250. The BWC is very light touch and low cost and may therefore be more cost effective than these other interventions.

The format of the BWC appeals to teachers and is easy to use, with teachers reporting very few problems related to using the website. A large majority of teachers who used the BWC report that they would like to use it again.

The effectiveness of the BWC could likely be improved by including more words that were unknown to students at the beginning of the intervention and by increasing the amount of time that teachers use the BWC lessons.

## Tables & Figures

**Table 1: Characteristics of schools by treatment (T) and control (C) assignment and sample**

<b>Variable</b>	<b>All</b>	<b>T</b>	<b>C</b>	<b><i>p</i>-value for difference between T and C</b>
<b>All randomized schools</b>				
Percent private school	17.0	14.8	19.2	0.676
Percent Title 1	63.4	66.7	60.0	0.626
Mean percent receiving free lunch within schools	54.1	59.8	48.0	0.239
Total number of schools	53	27	26	
<b>Schools in which students were assessed</b>				
Percent private school	19.1	15.4	23.8	0.476
Percent Title 1	63.8	69.2	57.1	0.402
Mean percent receiving free lunch within schools	55.0	61.8	46.5	0.150
Total number of schools	47	26	21	



**Table 2: Characteristics of students in treatment (T) and control (C) students who took the 17-week BWC Assessment**

Variable	All assessed students				Students with valid test scores			
	Total sample	T	C	<i>p</i>	Total sample	T	C	<i>p</i>
<b>Percent female</b>	50.4	49.6	51.4	0.553	51.6	50.6	52.8	0.454
<b>Percent kindergarten</b>	45.5	48.0	42.3	0.698	47.2	49.2	44.6	0.756
<b>Age in years</b>	5.73	5.79	5.66	0.494	5.76	5.81	5.68	0.482
<b>Percent OESL</b>	8.8	7.1	11.0	0.499	8.2	7.0	9.7	0.625
<b>Percent OSN</b>	6.5	6.8	6.0	0.779	5.2	5.8	4.4	0.529
<b>Number of students</b>	818	454	364		772	431	341	

Notes: OESL is the assessor’s observation of whether the student spoke English as a second language. OSN is the assessor’s observation of whether the student had special needs. The *p*-value for the difference between the treatment group and the control group is the *p*-value of the regression coefficient on treatment in a regression prediction each variable in the first column and with standard errors corrected for clustering in schools.

**Table 3: Characteristics of students in treatment (T) and control (C) schools who took the 25-week BWC Assessment**

Variable	All assessed students				Students with valid test scores			
	Total sample	T	C	<i>p</i>	Total sample	T	C	<i>p</i>
<b>Percent female</b>	52.7	51.5	54.3	0.381	52.9	51.5	54.7	0.328
<b>Percent kindergarten</b>	45.3	46.7	43.4	0.824	45.4	47.0	43.4	0.808
<b>Age in years</b>	5.67	5.73	5.60	0.501	5.68	5.73	5.60	0.501
<b>Percent OESL</b>	9.5	8.3	10.9	0.702	9.2	7.8	10.9	0.631
<b>Percent OSN</b>	4.3	5.4	3.0	0.297	4.4	5.4	3.0	0.294
<b>Student observations</b>	603	336	267		597	332	265	

Notes: OESL is the assessor's observation of whether the student spoke English as a second language. OSN is the assessor's observation of whether the student had special needs. The *p*-value for the difference between the treatment group and the control group is the *p*-value for the regression coefficient on treatment in a regression prediction each variable in the first column and with standard errors corrected for clustering in schools.

**Table 4: Intent-to-treat estimates of the effect of access to the BWC on valid BWC Assessment scores**

	<b>ITT-C coefficient</b>	<b>ITT-T coefficient</b>
<i>17-week BWC Assessment</i>		
Treatment	1.319*** (.42)	1.194*** (.43)
Effect size	.284	.256
N of students	862	862
<i>25-week BWC assessment</i>		
Treatment	1.242*** (.43)	1.071 ** (.43)
Effect size	.267	.229
N of students	663	663

Notes: Standard errors are in parentheses and are corrected for clustering at the school level. Significance levels: \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. All models include the following covariates: state fixed effects, whether the student is in kindergarten, student gender, whether the school is private, whether the school is Title 1, percent of school eligible for free or reduced-price lunch, observed English as a second language, and observed special needs status. The effect size is the ratio of the treatment coefficient to the control group's standard deviation. The number of students is the number who we assessed plus the number of students in schools that dropped out of the intervention after randomization. ITT-C estimates impute test scores for students in schools that dropped out with the mean test score of control group students. ITT-T imputes scores using the mean score of student sin the treatment group.

**Table 5: ITT-C estimates of the effect of access to the BWC on valid BWC Assessment scores for various samples**

	<b>ITT-C coefficient</b>
<i>17-week BWC Assessment</i>	
Treatment	1.319*** (.420)
Effect size	.284
Observations	862
<i>25-week BWC Assessment</i>	
Treatment	1.242*** (.430)
Effect size	.267
Observations	663
<i>17-week BWC Assessment for students assessed at week 25</i>	
Treatment	1.509*** (.420)
Effect size	.337
Observations	663

Notes: Standard errors are in parentheses and are corrected for clustering at the school level. Significance levels: \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. All models include the following covariates: state fixed effects, whether the student is in kindergarten, student gender, whether the school is private, whether the school is Title 1, percent of school eligible for free or reduced-price lunch, observed English as a second language, and observed special needs status. The effect size is the ratio of the treatment coefficient to the control group's standard deviation. ITT-C estimates impute test scores for students in schools that dropped out with the mean test score of control group students.

**Table 6: Treatment-on-treated estimates of the effect of access to the BWC on valid BWC Assessment scores**

<b>17- week BWC Assessment</b>	<b>Coefficient</b>
Treatment	1.638*** (.45)
Effect size	.320
N of observations	772
<b>25-week BWC Assessment</b>	
Treatment	1.561** (.51)
Effect size	.305
N of students	597

Notes: Standard errors are in parentheses and are corrected for clustering at the school level. Significance levels: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . All models include the following covariates: state fixed effects, whether the student is in kindergarten, student gender, whether the school is private, whether the school is Title 1, percent of school eligible for free or reduced-price lunch, observed English as a second language, and observed special needs status. The effect size is the ratio of the treatment coefficient to the control group's standard deviation.

**Table 7: Intent-to-treat (ITT) and treatment-on-treated (TOT) effect of access to the BWC on age-adjusted PPVT scores at week 25 for students with valid scores**

	<b>ITT-C</b>	<b>ITT-T</b>	<b>TOT</b>
<b>Treatment</b>	1.891 (1.25)	1.815 (1.25)	2.44 (1.55)
<b>Effect Size</b>	.135	.129	0.158
<b>Number of students</b>	651	651	595

Notes: Standard errors are in parentheses. Estimates are corrected for clustering at the school level. All models include the following covariates: state fixed effects, whether the student is in kindergarten, student gender, whether the school is private, whether the school is Title 1, percent of school eligible for free or reduced-price lunch, observed English as a second language, and observed special needs status. The effect size is the ratio of the treatment coefficient to the control group's standard deviation. ITT-C estimates impute test scores for students in schools that dropped out with the mean test score of control group students. ITT-T imputes scores using the mean score of students in the treatment group.

**Table 8: Effect of access to the BWC on BWC Assessment scores for various subgroups**

Variable	17-Week BWC Score		25-Week BWC Score	
	Treatment Coefficient	Interaction Coefficient	Treatment Coefficient	Interaction Coefficient
Female	1.181** (.519)	0.261 (.531)	0.431 (.535)	1.487** (.666)
Kindergarten	0.867 (.562)	0.949 (.898)	0.662 (.725)	1.23 (1.006)
Private school	1.54*** (.463)	-1.137 (1.062)	1.259*** (.453)	-0.084 (1.493)
Title 1 school	0.841 (.758)	0.745 (.972)	0.986 (.877)	0.399 (1.091)
OESL	1.307*** (.456)	0.130 (1.164)	1.461*** (.457)	-2.154 (1.631)
OSN	1.404*** (.451)	-0.857 (1.202)	1.189*** (.446)	1.538 (1.391)
<b>Number of students</b>	862		663	

Notes: Standard errors are in parentheses. Estimates are corrected for clustering at the school level. OESL is the assessor's observation of whether the student spoke English as a second language. OSN is the assessor's observation of whether the student had special needs. The effect size is the ratio of the treatment coefficient to the control group's standard deviation. These are estimated from the following model:

$$F_i = \hat{\beta}_0 + \hat{\beta}_1 T_i + \epsilon_i$$

where F equals 1 if the student has the target characteristic (female, kindergarten) and 0 otherwise, and T is equal to one if the student was in the treatment group and 0 if the student was in the control group. e school-level.

The other equations are:

$$Age_i = \hat{\beta}_0 + \hat{\beta}_1 T_i + \epsilon_i, \text{ Age is the student's age}$$

$$ESL_i = \hat{\beta}_0 + \hat{\beta}_1 T_i + \epsilon_i, \text{ ESL is whether the student was observed as being ESL}$$

$$SN_i = \hat{\beta}_0 + \hat{\beta}_1 T_i + \epsilon_i, \text{ SN is whether the student was observed as being special need}$$

All standard errors clustered at the school level, and all p-values test the null hypothesis that there is no mean difference between the control and treatment group or  $\beta_1 = 0$ .

**Table 9: Treatment teacher logins to the BWC website by date**

	<b>To December 12, 2017</b>	<b>To January 31, 2018</b>	<b>To April 3, 2018</b>	<b>To May 7, 2018</b>
Median logins this period	13	22	39	45
Cumulative mean logins per teacher	14.6	30.1	50.6	60.0
Maximum potential school days this period	20	25	44	23
Cumulative maximum potential school days	20	45	89	112
Mean login per potential school day this period	.73	.67	.57	.54

Notes: Login data provided by the BWC staff. Potential school days are the number of weekdays in the login period even though school may not have been in session every weekday.



**Table 10: Treatment teacher reports of their use of the components of the BWC program**

<b>In a typical week, how many <u>days per week</u> did you use the following BWC program components?</b>	<b>Percent who used Word of the Day</b>	<b>Percent who used the book</b>	<b>Percent who used the song</b>	<b>Percent who used the dance</b>	<b>Percent who used the review</b>
Never	0	15.4	4.0	9.5	14.3
Less than once per week	4.8	15.4	6.0	21.4	23.8
1–2 times per week	38.1	53.8	25.0	57.1	57.1
3–4 times per week	21.4	12.8	5.0	7.1	4.8
Every day	35.7	2.6	2.0	4.8	0
Number of teachers responding	42	39	42	42	42

Notes: Data are from the survey of teachers who participated in the BWC evaluation

**Table 11: Reasons treatment teachers did not use the BWC**

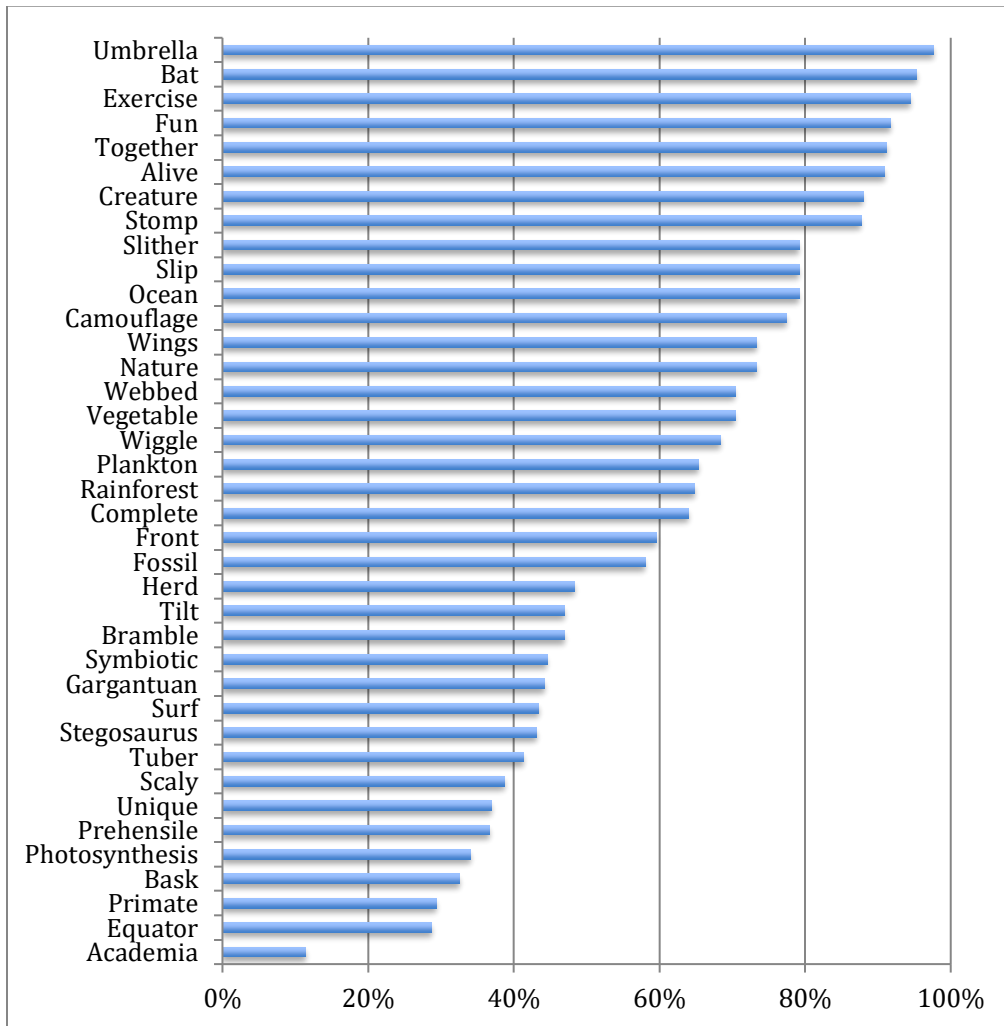
<b>On days that you did not use the BWC, what was the main reason for not using it?</b>	<b>Percent saying this was the main reason (n=35)</b>
Too many other required activities	65.7
Thought students would learn more from something else	8.6
Not interested in using it	5.7
Students not interested in using it	2.9
Technological issues	11.4
I forgot	28.6

Notes: Data are from the survey of teachers who participated in the BWC evaluation. These responses are from treatment group teachers only (n=35). Teachers could list more than one reason for not using the BWC.

**Table 12: Treatment teacher views on the content of the BWC**

	<b>Percent of teachers</b>
<i>Vocabulary is:</i>	
A little too easy	3.0
About right	48.5
A little too hard	36.4
Much too hard	12.1
Number responding	33
<i>Length:</i>	
A little too long	14.7
About right	82.4
A little too short	2.9
Number responding	34
<i>How appealing?</i>	
Very appealing	35.3
Appealing	58.8
Not so appealing	5.9
Number responding	34

**Figure 1: Percent of control group students correctly identifying each word on the BWC Assessment at Week 17**



## Appendices

### Appendix A: Sample Email to Principals to Obtain Teacher Participation Approval

From: [REDACTED]  
Subject: RE: Big Word Club Letter of Participation  
Date: October 13, 2017 at 1:35:18 PM MDT  
To: 'Gillian Martin' <gillian@bigwordclub.com>

Thanks! I approve!

[REDACTED]  
[REDACTED]  
[REDACTED]  
[REDACTED]

Elementary Principal

Fax: [REDACTED]

The information in this transmission may contain privileged and confidential information and is intended only for the use of the person(s) named above. If you are not the intended recipient, or an employee or agent responsible for delivering this message to the intended recipient, any review, dissemination, distribution or duplication of this communication is strictly prohibited. If you are not the intended recipient, please contact the sender immediately by reply email and destroy all copies of the original message.

From: Gillian Martin [mailto:gillian@bigwordclub.com]  
Sent: Friday, October 13, 2017 1:26 PM  
To: [REDACTED]  
Cc: [REDACTED] ; Shane DeRolf  
Subject: Big Word Club Letter of Participation  
Hi Melissa,

Happy Friday!

We're reaching out to confirm your interest in being part of the Big Word Club study being funded by MIT. Your classroom was selected out of over 500 applicants and we'd love for you to be part of the study. If you're no longer interested, would you please let us know so we can let another class in?

If you are still interested, and we hope you are, we need your principal to reply to this email with the words "I agree" typed in her response.

Please understand that your classroom will not be enrolled until we receive Principal [REDACTED] "I agree" email.

Thank you again for being part of this important (and fun) study!

All the best,

Shane

██████████, This is a letter of agreement for classrooms in ██████████ Elementary School to participate in the evaluation of Big Word Club. One of your teachers, ██████████, has volunteered to have her classroom participate. The average amount of time required to implement Big Word Club is only four minutes a day. A complete description of Big Word Club can be found at <https://bigwordclub.com>. The evaluation of Big Word Club is intended to determine the extent to which the program increases the vocabularies of children in preschool and kindergarten. The evaluation is being funded by the Abdul Latif Jameel Poverty Action Lab at the Massachusetts Institute of Technology and is being conducted by Professor Phil Oreopoulos of the University of Toronto and Professors Ariel Kalil and Susan E. Mayer at the University of Chicago. The evaluation will consist of comparing scores on an assessment of vocabulary words from Big Word Club and other age-appropriate vocabulary words between classrooms in schools that are in the “treatment” group and therefore receive access to Big Word Club and a “control” group of schools that do not receive Big Word Club in the 2017-18 school year. Teacher Responsibilities: The evaluation will require nothing of teachers other than that they use Big Word Club in their classrooms for as much or as little as they would like and that they participate in a brief survey about their experience with Big Word Club if they choose to do so. In appreciation for your school’s participation in the evaluation, the entire school will have free access to Big Word Club for the 2018-19 school year. In agreeing that classrooms in your school may participate in the evaluation, you agree to the following:

- ██████████ Elementary School may be assigned to the “treatment” group in which classrooms receive the Big Word Club for the 2017-18 school year or to the “control” group in which the classrooms will not receive the Big Word Club for the 2017-18 school year. Each school has about a 50% chance of being in the treatment group.
  - Children in classrooms participating in the evaluation will be assessed twice by an outside assessor; once after 18 weeks and once after 27 weeks.
  - Assessments will require that children (in groups of approximately five students) be taken out of the classroom for about 20 minutes for each assessment.
  - The school will provide a quiet private space where the assessments can take place.
  - If required, ██████████ Elementary School is responsible for obtaining parental consent for the assessments in accordance with the policies of your school district.

To provide your agreement that classrooms in ██████████ Elementary School have permission to participate, please hit reply all and type in “I agree” to this email.

If you have any questions about the evaluation or about Big Word Club, please contact me by return email and I (or a member of the research team) will be happy to get back to you.

Warm regards,

Gillian Martin



Gillian Martin / Project Manager

[gillian@bigwordclub.com](mailto:gillian@bigwordclub.com)

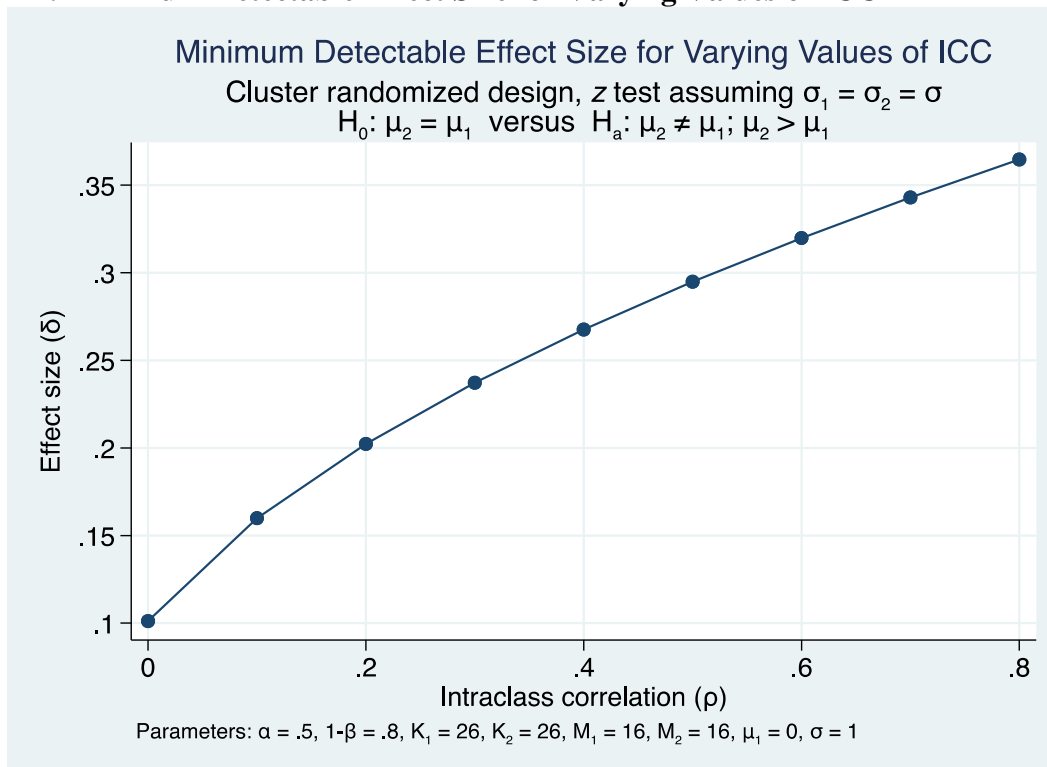
(303) 550.7700

[bigwordclub.com](http://bigwordclub.com)

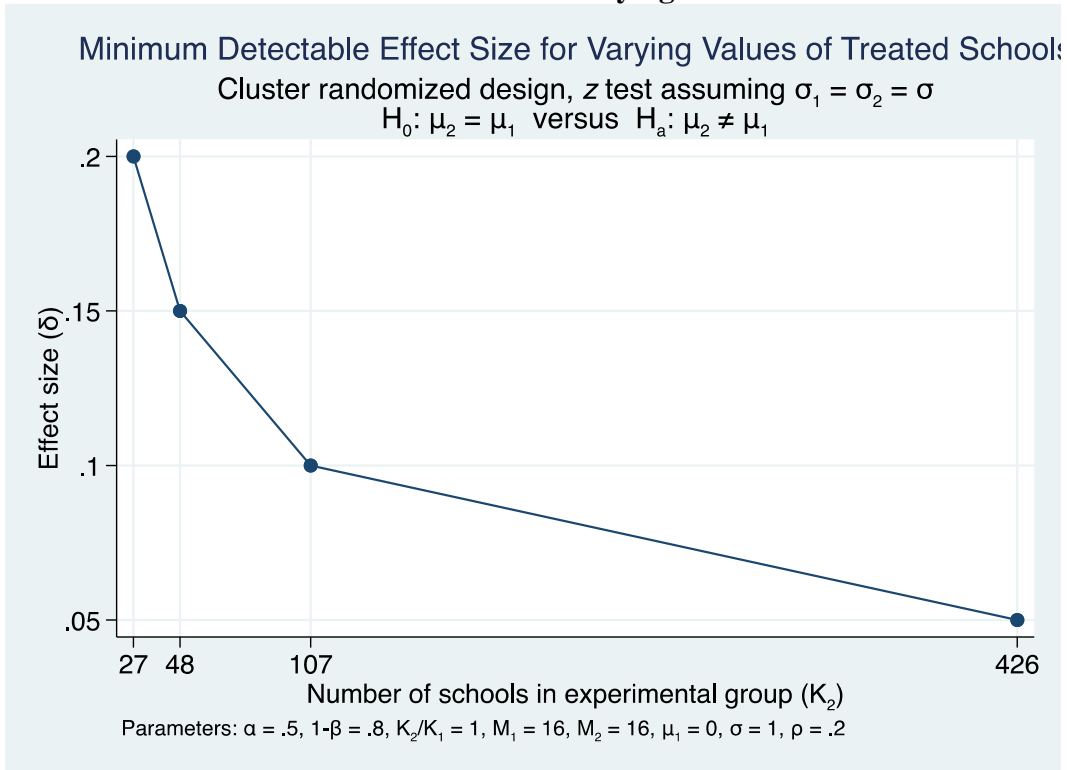
## Appendix B: Power Calculations and Sensitivity Analysis

**Power calculation.** We based the power calculations on the assumption that we would have 52 schools with 16 students each (for a total of 832 students) and half the schools would be assigned to the treatment. We also assumed an intraclass correlation of 0.2 and set power to 80% and significance level to 5%. Given these assumptions, the minimum detectable effect size (MDES) was 0.202. In other words, students in treatment schools would have to score more than 0.20 standard deviations higher than control students for us to detect treatment effects at the assumed power and significance level. If intraclass correlation is higher, say 0.4, the MDES goes up to 0.268. Figure A1 shows the predicted changes in the MDES under various values of the interclass correlation. Figure A2 shows that an increase in the number of students per school does not improve the MDES by much. If we recruit 100 students per school, MDSE goes down to 0.185 (given power = 0.8, alpha = 5%, ICC = 0.2, and total schools = 52).

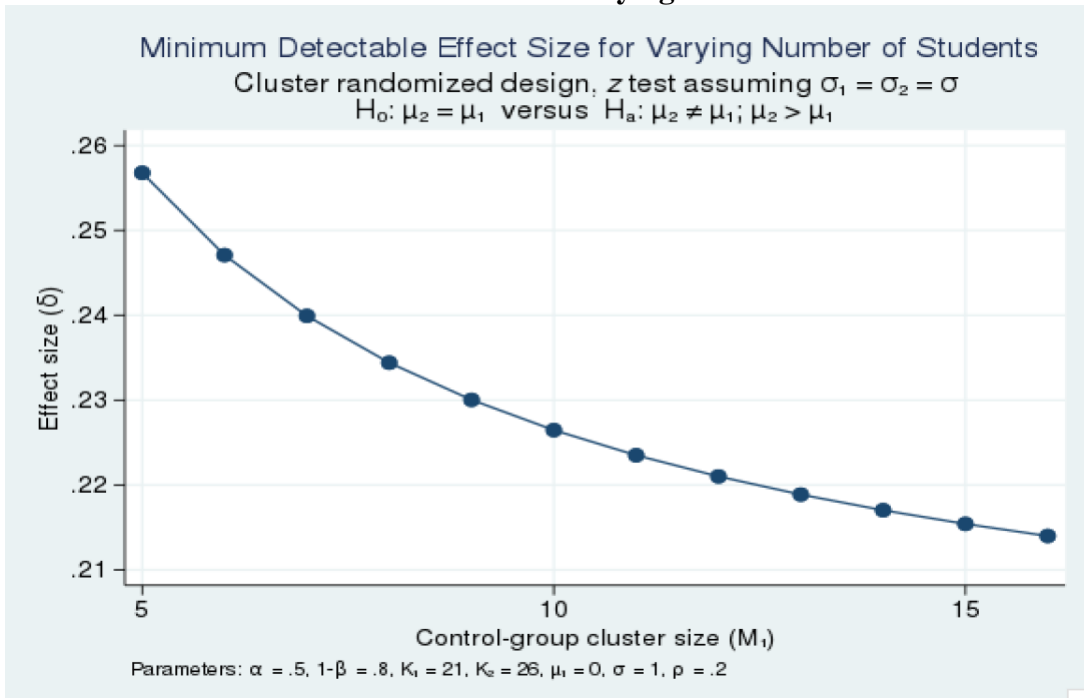
**Figure A1: Minimum Detectable Effect Size for Varying Values of ICC**



**Figure A2: Minimum Detectable Effect Size for Varying Values of Treated Schools**



**Figure A3: Minimum Detectable Effect Size for Varying Number of Students**





**Table A1: ITT estimates of the effect of access to the BWC on the BWC Assessment for all students with valid scores controlling only state and student age for students with valid test scores**

	ITT-C	ITT-T
<i>17-week BWC Assessment</i>		
Treatment	0.906 (4.646)	0.700 (4.698)
Effect size	.195	.149
N	862	862
<i>25-week BWC Assessment</i>		
Treatment	0.799 (4.645)	.619 (4.689)
Effect size	.172	.132
N	663	663
<i>25-week PPVT Assessment</i>		
Treatment	.329 (14.304)	.250 (14.706)
Effect size	.023	.017
N	651	651

Notes: Standard errors are clustered at the school level. Significance levels: \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. All models include age and state as covariates. The effect size is the ratio of the treatment coefficient to the control group's standard deviation.

**Table A2: TOT estimates of the effect of access to the BWC on the BWC Assessment controlling only state and student age for students with valid test scores**

<b>17-week BWC Assessment</b>	<b>Estimate</b>
Treatment	1.009 (5.122)
Effect size	.197
N	772
<b>25-week BWC Assessment</b>	
Treatment	.886 (5.121)
Effect size	.173
N	597
<b>25-week PPVT Assessment</b>	
Treatment	.183 (14.077)
Effect size	.013
N	587

Notes: Standard errors are clustered at the school level. Significance levels: \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. All models include age and state as covariates. The effect size is the ratio of the treatment coefficient to the control group's standard deviation.

## Appendix C: The BWC Assessment

The evaluation covered only words included in the first 16 weeks of the BWC. Over this period, children would in principle be exposed to 80 words in the BWC program if teachers used the program five days a week every week. In practice, because of holidays and other periods when school is closed the actual potential number of words that students would be exposed to is less than this; that number varies by school. In addition, not all of the words that children are exposed to in the BWC are new to all children. As we discuss below, many of the words in the BWC are already familiar to most preschool and kindergarten children.

The TOT estimate is the ratio of two experimental ITT effects: the ITT effect on the outcome of interest ( $Y$ ) divided by the ITT effect on participation rates in the intervention being studied. This method recovers the TOT if assignment to treatment has no effect on outcomes for subjects who do not participate. In this case we define participation in the treatment as having logged on to the BWC website at least once during the treatment period.

The first is a meta-analysis of 67 interventions intended to improve the receptive and expressive vocabulary of preschool and kindergarten children (Marulis & Neuman 2010). It found an average effect size of  $g = 0.88$ ,  $p < .001$  across the studies.<sup>5</sup> Vocabulary gains on standardized assessments were significantly lower ( $g = 0.71$ ,  $p < .01$ ) than those on author-created measures ( $g = 1.21$ ,  $p < .01$ ). However, few of the studies included in the meta-analysis are comparable to the BWC. The meta-analysis included studies that contained training, intervention, or specific teaching techniques to increase word learning. A large number of the interventions were not intended to be used in classrooms but rather were intended to answer theoretical or practical questions about how children learn vocabulary. Experimenters delivered twenty-five of the interventions. When experimenters deliver the intervention it assures extremely high fidelity that is unlikely to be replicated in a real classroom setting. Some of the interventions were for children with specific needs, including some for sign language and English as a second language. A third of the studies were done more than 20 years ago, which raises questions about the similarity of classroom settings over such a long time period. The vast majority of the interventions were based on various forms of storybook reading,

The second meta-analysis is more relevant in that the analysis included only vocabulary instruction programs intended for classroom use. Studies that only used repeated readings, read-alouds, or independent reading were also excluded from this review unless the intervention contained an instructional method for teaching vocabulary. But the meta-analysis covered 37 vocabulary interventions evaluated using RCTs in grades pre-K to 12 (Elleman et al., 2009). If vocabulary interventions work differently for older children, not all of the studies in the meta-analysis are relevant. In addition, the outcome in all of these studies was reading comprehension at the passage level (four or more sentences), which is quite different from receptive vocabulary and not especially relevant for preschool and kindergarten children. This meta-analysis showed that on average vocabulary instruction programs increased students' text comprehension when the assessment measure was designed to test the words in the intervention program ( $d = 0.50$ ), but the effect size was smaller when the assessment was a standardized measure of vocabulary or

---

<sup>5</sup> In this meta-analysis all effect sizes are estimated using Hedges's  $g$  coefficient, a more conservative form of the Cohen's  $d$  effect size estimate (Lipsey & Wilson, 1993).

related skills ( $d = 0.10$ ).

The programs included in both these meta-analyses were delivered to children with no mental or physical limitations and in English. The classroom-based interventions that were included in both meta-analyses were mainly curriculum-based interventions often requiring extensive teacher training. Many of the interventions were delivered individually or in groups of fewer than five children. The need for teacher training and the need to deliver an intervention in small groups can raise the cost of an intervention greatly.

Because of the many differences between the studies included in the meta-analysis and the BWC, they provide little context for the effectiveness of the BWC.

**Please click on the following link to access the full BWC Assessment and Score Sheet:**

**<https://sites.google.com/view/bip-bwc-online-appendix/home>**

## References

- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, 91(434), 444–455.
- Apthorp, H., Randel, B., Cherasaro, T., Clark, T., McKeown, M., & Beck, I. (2012). Effects of a Supplemental Vocabulary Program on Word Knowledge and Passage Comprehension. *Journal of Research on Educational Effectiveness*, 5(2), 160–188. doi: 10.1080/19345747.2012.660240.
- Beck, I. L., & McKeown, M. G. (2007). Increasing Young Low-Income Children's Oral Vocabulary Repertoires through Rich and Focused Instruction. *Elementary School Journal*, 107(3), 251–273.
- Biemiller, A. (2005). Size and sequence in vocabulary development: Implications for choosing words for primary grade vocabulary instruction. In A. Hiebert & M. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 223–242). Mahwah, NJ: Erlbaum.
- Bloom, H. (1984). Accounting for No-Shows in Experimental Evaluation Designs. *Evaluation Review* 8(2), 225-246.
- Christie, J., Roskos, K., Vukelich, C., & Han, M. (2003). The effects of a well-designed literacy program on young children's language and literacy development. In F. Lamb-Parker, J. Hagen, R. Robinson, & H. Rhee (Eds.), *The first eight years. Pathways to the future: Implications for research, policy, and practice*. Proceedings of the Head Start National Research Conference (pp. 447–478). New York: Mailman School of Public Health, Columbia University.
- Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, 33(6), 934–945.
- Dickenson, D. K., & Porche, M. V. (2011.) Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development*, 82(3), 870-886.
- Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The Impact of Vocabulary Instruction on Passage-Level Comprehension of School-Age Children: A Meta-Analysis. *Journal of Research on Educational Effectiveness*, 2(1), 1–44. doi: 10.1080/19345740802539200.
- Farkas, G., & Beron K. (2004). The detailed trajectory of oral vocabulary knowledge: Differences by class and race. *Social Science Research*, 33(3), 464–497.
- Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., & Martin, A. J. (1999). Phonological Short-term Memory and Vocabulary Development: Further Evidence on the Nature of the Relationship. *Applied Cognitive Psychology*, 13(1): 65–77.
- Goodson, B., Wolf, A. Bell, S., Turner, H., Finney, P. B. (2010). *The Effectiveness of a Program to Accelerate Vocabulary Development in Kindergarten (VOCAB)*. NCEE 2010-4014. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education.
- Marulis, Loren M., & Susan B. Neuman. (2010). "The Effects of Vocabulary Intervention on

- Young Children's Word Learning: A Meta-Analysis," *Review of Educational Research*, 80 (3): 300–335. doi: 10.3102/0034654310377087.
- Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015). 24-Month-Old Children With Larger Oral Vocabularies Display Greater Academic and Behavioral Functioning at Kindergarten Entry. *Child Development*, 86(5), 1351–1370. doi: 10.1111/cdev.12398.
- National Early Literacy Panel. (2009). *Developing Early Literacy: Report of the National Early Literacy Panel: A Scientific Synthesis of Early Literacy Development and Implications for Intervention*. Washington, DC: National Institute for Literacy, National Center for Family Literacy.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Bethesda, MD: National Institute of Child Health and Human Development. ERIC Document Reproduction Service No. ED444127.
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98(3), 554–566.
- Preschool Curriculum Evaluation Research (PCER) Consortium. (2008). *Creative Curriculum with Ladders to Literacy*: University of New Hampshire. In *Effects of preschool curriculum programs on school readiness* (pp. 65–73). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S.
- Roskos, K., Ergul, C., Bryan, T., Burstein, K., Christie, J., & Han, M. (2008). "Who's learning what words and how fast? Preschoolers' vocabulary growth in an early literacy program. *Journal of Research in Childhood Education* 22(3) 275-290.
- Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development*, 83(2), 508–525.
- Scarborough, H. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. B. Neuman & D. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). New York: Guilford.
- Spira, E. G., Bracken, S. S., & Fischel, J. E. (2005). "Predicting improvement after first-grade reading difficulties: The effects of oral language, emergent literacy, and behavior skills." *Developmental Psychology*, 41(1), 225–234.
- Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Stahl, S. A., & Nagy, W. E. (2006). *Teaching word meanings*. Mahwah, NJ: Erlbaum.
- Storch, S. A., & Whitehurst, G. J. (2002). "Oral language and code-related precursors to reading: Evidence from a longitudinal structural model." *Developmental Psychology*, 38(6), 934–947.
- Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W., and &, M. (2007). The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills, word identification skills, and reading comprehension by children with reading disabilities. *Journal of Speech, Language, and Hearing Research*, 50(4), 1093–1109.



BEHAVIORAL INSIGHTS  
AND PARENTING LAB  
**THE UNIVERSITY OF CHICAGO**