WORKING PAPER · NO. 2024-15

Making a Song and Dance About It: The Effectiveness of Teaching Children Vocabulary with Animated Music Videos

Ariel Kalil, Susan Mayer, Philip Oreopoulos, and Rohen Shah

FEBRUARY 2024

Becker Friedman Institute
FOR ECONOMICS AT UCHICAGO

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

MAKING A SONG AND DANCE ABOUT IT: THE EFFECTIVENESS OF TEACHING CHILDREN VOCABULARY WITH ANIMATED MUSIC VIDEOS

Ariel Kalil
Susan Mayer
Philip Oreopoulos
Rohen Shah

This work was supported by J-PAL. This study was registered on the AEA RCT Registry (AEARCTR-0002631) and received approval from the Social and Behavioral Sciences IRB from the University of Chicago (IRB17-1609). We are grateful to the staff at the Behavioral Insights and Parenting Lab, led by Michelle Park Michelini, for invaluable effort in implementing this intervention, and Harkirat Kaur for excellent research assistance. 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. We also thank participants at the Advances with Field Experiments and AEFP conferences for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Making a Song and Dance About It: The Effectiveness of Teaching Children Vocabulary with Animated Music Videos

Ariel Kalil, Susan Mayer, Philip Oreopoulos, and Rohen Shah

February 2024
JEL No. D91,I20,J10,O15

ABSTRACT

Programs that engage young children in movement and song to help them learn are popular but experimental evidence on their impact is sparse. We use an RCT to evaluate the effectiveness of Big Word Club (BWC), a classroom program that uses music and dance videos for 3-5 minutes per day to increase vocabulary. We conducted a field experiment with 818 preschool and kindergarten students in 47 schools in three U.S. states. We find that treated students scored higher on a test of words targeted by the program (0.30 SD) after four months of use and this effect persisted for two months.

Ariel Kalil
Harris School of Public Policy
University of Chicago
1307 E. 60th Street
Chicago, IL 60637
akalil@uchicago.edu

Susan Mayer
Harris School of Public Policy
University of Chicago
1307 E. 60th Street
Chicago, IL 60637
smayer@uchicago.edu

Philip Oreopoulos
Department of Economics
University of Toronto
150 St. George Street
Toronto, ON M5S 3G7
and NBER
philip.oreopoulos@utoronto.ca

Rohen Shah
The University of Chicago
Harris School of Public Policy
1307 E. 60th St
Chicago, IL 60637
United States
shahr@uchicago.edu

A randomized controlled trials registry entry is available at AEARCTR-0002631

1 Introduction

Vocabulary acquisition is a crucial part of developing literacy skills (Cunningham & Stanovich, 1997; Stahl & Nagy, 2007), and measuring the effectiveness of various types of vocabulary-improvement programs can help parents, schools, and states decide how to invest in different approaches. Children's vocabulary skills are often the target of multisensory programs that aim to educate with music, dance, song, and rhymes using animation, live action, or puppetry. Well-known children's television programs such as Sesame Street, the Electric Company, and Schoolhouse Rock aim to teach children vocabulary, math, creativity, and social-emotional skills with these approaches. We study this approach of providing education with engaging content, or what Banerjee et al. (2019) refer to as *edutainment*. In this paper, we use an RCT to evaluate the effectiveness of such a vocabulary program, *Big Word Club* (BWC).

BWC is a series of animated music videos intended to help children learn one new word per day. Cheerful videos show animal characters dancing and playing along to an original song with clever, catchy Randy Newman-esque lyrics and music. The music is arranged with minimal instrumentation (think banjo, clarinet, guitar) so that the lyrics are the center of attention. For example, the video for "The Symbiotic Song" portrays a hippo with a bird riding along on the hippo's back. The bird munches small bugs on the hippo's back and alerts the hippo to a lion lying in wait on the savannah. The animals sing "we're symbiotic, you and me; we're symbiotic it's a great big we…we're symbiotic helping each other through, it's us together it's me and you…what's good for one is good for two."[2] The program can be used both by parents at home and teachers in school and is designed for children in preschool and elementary school (ages 4-

[2] More BWC videos can be seen here: https://www.youtube.com/playlist?list=PLbnT4BbeKNIdy-Ch0dYd8dT_A_tXAbwAj

10). Because the program only takes 3-5 minutes per day, teachers have considerable flexibility in how they incorporate the program into their class.

We evaluate the effectiveness of the BWC program at improving children's vocabulary skills using a cluster randomized RCT where teachers in treatment schools received access to the BWC program. The field experiment was conducted with 818 pre-K and kindergarten students in 47 schools across 3 U.S. states. We designed a vocabulary assessment of words covered in the first four months of the BWC program. Students took this assessment four months after the treatment began and again two months later to test treatment persistence. To test the impact of the treatment on vocabulary skills more broadly on words not included in the BWC videos students took a standardized vocabulary assessment (PPVT) six months after the treatment began.

We estimate the Intent-to-Treat (ITT) effect, which is the policy-relevant estimate for a program such as this that is intended to be implemented at scale. We find that treatment students scored significantly higher (0.30 SD) on the assessment of BWC word knowledge than control students. Further, this effect persisted 2 months later in the follow-up assessment. We find no statistically significant difference in PPVT scores between the treatment and control groups, implying both that the treatment effect did not come at the expense of other vocabulary words, and that it did not encourage language development beyond the words targeted by the program.

Our study makes several contributions to understanding skill development using educational media. First, the broader economic literature on "edutainment" mainly studies the impact of watching media that aims to provide information with entertaining content in the service of attitude and behavior change. These studies are interested in the relative effectiveness of "edutainment" compared to traditional educational content. The delivery mechanism for

"edutainment" is commercial television, typically a soap-opera, tele-novella, or "game-show" format. Outcomes of interest in these studies have centered on attitudes, beliefs and behavioral choice among young adults on topics such as domestic violence, entrepreneurship, risky sexual behavior, and financial decision-making (Banerjee et al., 2015, 2019; Barsoum et al., 2022; Berg & Zia, 2017; Coville et al., 2019; Ravallion et al., 2015). Most such studies have been conducted in developing country contexts (but see Kearney & Levine (2015) for a test of American mass media on US teenagers' fertility decisions).

More relevant to the present investigation is the very small number of experimental studies that have examined the impact of children's exposure to educational media on child skill development. Virtually all of these studies evaluate content from Sesame Street or similar media. Given its enormous popularity, it is surprising that Sesame Street itself has been subject to only one well-designed and well-powered experimental evaluation in the US at the time of its initial broadcast (Bogatz & Ball, 1971). It has similarly been the focus of only one long-run follow up using quasi-experimental methods to assess outcomes at the population level (Kearney and Levine, 2019). And, although Sesame Street is now a global phenomenon, we could only identify two experimental evaluations of its impact on child skills outside the United States, in Mexico and Jordan, respectively. The study in Mexico was conducted in the early 1970s, whereas the study in Jordan was conducted very recently with children displaced by conflict in the MENA region. Other recent relevant interventions bundle educational media programming with teacher-led analog games, making it hard to determine the unique impact of the educational media (Penuel et al. 2012).

The present study adds to this small body of work by testing whether a very light touch classroom intervention with similar types of engaging educational media can improve and

sustain young children's vocabulary. This program, the *Big Word Club*, has several unique features. First, it takes only 5 minutes per day and requires no extra programming. It also requires no teacher training and can be used in class or at home. With respect to classroom interventions in the US, it is hard to imagine schools adopting an educational media curriculum that takes 30 minutes per day every day, as other interventions have tested. Thus, from a contemporary educational policy perspective the *Big Word Club* is highly relevant. Moreover, existing randomized control trials either include only low-income preschoolers, were conducted more than 50 years ago, lack follow-up beyond the treatment period, or were conducted outside the US. These features make it hard to draw inferences relevant to contemporary learning and educational policy in the US.

In contrast, our study uses current data from US children sampled from public schools serving families with a range of socioeconomic status. We measure children's vocabulary skill at the end of the intervention and in a short-term follow-up. This paper proceeds as follows. In section 2, we discuss the relevant literature in more detail and the Big Word Club program. Section 3 describes our experimental design and Section 4 presents the results. Section 5 discusses the study's findings and limitations.

2 Background

2.1 Importance of Vocabulary

Vocabulary is central to oral language development, reading comprehension, and development of domain-specific knowledge (Beck & McKeown, 2007; Cunningham & Stanovich, 1997; Scarborough et al., 2009; Stahl & Nagy, 2007). Research suggests that there are

large differences in vocabulary knowledge between children from affluent and low-income backgrounds by the age of three (Hart & Risley, 1995; Suskind et al., 2015), leading educators to strongly advocate for vocabulary learning in early childhood education.

2.2 Educational Media and Child Vocabulary Skill Development

The best-known approach to teaching children through educational media is the beloved and influential Sesame Street, which was created in 1969 to harness the power of television to help boost the literacy and numeracy skills of low-income 3-5 year old children who were not enrolled in preschool. Sesame Street was designed to be more than just entertainment; it was a deliberate effort to make learning fun. The show incorporates educational content, including letters, numbers, and basic life skills, into engaging and entertaining segments that use song, dance, and puppetry. The program was designed in collaboration with educators to align with young children's developmental needs and has evolved over time to include interactive and digital content (Fisch & Trulio, 2001). From 1969-2015 each episode of Sesame Street was 60 minutes long. In 2014 the episode length was reduced to 30 minutes.

The Educational Testing Service conducted the first experimental evaluations of Sesame Street in its earliest years. The more well-powered of the two evaluations it completed focused on about 630 disadvantaged at-home preschool children in the second year the program aired (Bogatz & Ball, 1971). For this evaluation, the researchers sampled low-income families in North Carolina and Los Angeles who had not had exposure to the program in the first year and, by virtue of the cost of cable as well as the cost of televisions that could receive cable, could not afford the technology necessary to view the program. In North Carolina, the researchers collaborated with the local cable company to randomly assign cable television access to families

free of charge on some blocks in the neighborhood. Other families were not given free cable television and served as controls. In Los Angeles, the researchers purchased the necessary technology for families in the treatment group to allow them access to viewing the program. Parents in all treatment groups were also encouraged to watch the show. Children were tested in their homes on a variety of outcomes. Results showed that children in the treatment group made significant gains in many areas including pre-reading skills and early math skills. For instance, treated children improved in their PPVT scores by as much as .36 standard deviations, which research suggests is equivalent to a full year of learning (Kearney & Levine, 2019, p. 323).

The only rigorous long-term evaluation of Sesame Street in the US is a quasi-experimental evaluation conducted by Kearney and Levine (2019). This study relied on random geographic variation in broadcast reception to study the effects of exposure to the program in its initial year (1969). Outcome data were collected from the census on broad measures of attainment including grade retention, educational attainment, and labor market outcomes. The results showed that children who enjoyed exposure to the show were 14 percent more likely to be in the appropriate grade for their age with larger point estimates for boys and non-white children.

Sesame Street is a global phenomenon. However, a meta-analysis of Sesame Street with 24 studies in 15 countries included only four RCTs, of which only two measured cognitive outcomes and only one was published in a peer reviewed academic journal (Mares & Pan, 2013). This study, conducted in Mexico using the Spanish version of Sesame Street, assigned control students to watch children's cartoons or non-educational television whereas treated students were assigned to watch 50 minutes of Sesame Street per day (Diaz-Guerrero & Holtzman, 1974). Data were collected from about 175 children who were sampled from preschools during the show's first telecast season. Preschool centers were randomly assigned to treatment or control group,

and children watched the program to which their center was assigned every afternoon for six months at the childcare centers. The results showed that treated four-year old children made very large gains in general knowledge (about two whole standard deviations) and about a 1.15 SD gain in letters and words as well as numbers. Treated children also made gains on a test of oral comprehension that was independent of the Sesame Street curriculum.

Sesame Street's most recent incarnation is *Ahlan Simsim* ("Welcome Sesame" in Arabic), a locally produced Arabic-language version of Sesame Street, airing across the Middle East and North Africa (MENA) and designed to help the skill development of young children affected by crisis and conflict in the region. Each season of *Ahlan Simsim* focuses on social-emotional development, including teaching specific emotion words and strategies for managing strong emotions. This program has recently been experimentally evaluated in Jordan in an intervention that plays the program to children in their preschool classrooms (Moran et al., 2023). 216 schools serving about 4300 pre-school age children participated in the RCT. The treatment, to which half of these schools were assigned, consists of watching 30-minute episodes of the program once per day in class for 12 weeks.

The evaluation *Ahlan Simsim* in preschool classrooms showed that children made significant improvements in expressive emotion recognition; i.e., their ability to state a specific emotion word when shown an illustration of a facial expression. Children made gains not only on the emotion words contained in the episodes for the treatment curriculum, but on other emotion words not in the episodes they watched as part of the intervention.

Finally, Penuel et al. (2012) use a cluster RCT to randomize 436 low income children in 80 preschool classrooms in New York and San Francisco to evaluate "Ready to Learn." This program is a teacher-delivered curricular supplement aimed at supporting learning among low

income children by providing them with 25 hours over 10 weeks (i.e., 30 minutes per day) of public media literacy programming from Sesame Street and similar programs. The Ready to Learn curriculum also included games and teacher-led non-video activities and required teacher training and in-class coaching. In this evaluation, Ready to Learn had large positive impacts ($+0.20 \leq d \leq +0.55$) on children's ability to recognize letters, sounds of letters and initial sounds of words, and children's concepts of story and print. But it is not possible to disentangle the effects of the educational media from the effects of games and other teacher-led activities.

It is not hard to imagine that an engaging media program that invites children to sing and dance while they learn vocabulary words will sustain attention, reduce boredom, and strengthen engagement in learning compared to traditional, rote learning methods like memorization. Qualitative work aimed at identifying drivers of the effects of popular vocabulary media programs has focused on two features: definition provision and attention-redirection (Neuman et al., 2019). Neuman et al. (2019) further uses eye-tracking technology to find that attention-redirection cues had a stronger correlation with successful word identification. As noted in Penuel et al. (2013) another positive feature of educational media interventions delivered in classrooms for young children is that they allow children to learn in social partnership with others and interacting with social partners while consuming this content can strengthen learning.

2.3 Big Word Club

Time-intensive programs, such as the daily 30-minute interventions described in the prior sections, may be hard to implement at scale (List, 2022). The opportunity cost of 30-60 minutes of daily class time is not negligible. An open question is whether a daily program that is less time-intensive and classroom-based would have a lasting impact on children's vocabulary skill.

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. It is intended for children in preschool to grade 5, with different classroom materials depending on the grade. The program takes approximately 5 minutes per day of class time.

The BWC words are "big" in the sense that many are not typical of the vocabulary of such young children. For example, the words for preschoolers include *gargantuan*, *primate*, and *equator*. Each week, the BWC provides classroom teachers with nine new videos based on that week's theme. The weekly videos include five that introduce the word for each day, one animated book, one music video, one dance video, and one review video – all of which include the five words for that week.

The BWC provides flexibility to teachers in that they can use the videos any time during the day. Each video is only 1-4 minutes long, so implementing the BWC is not costly in terms of classroom time. Many teachers report using the animated books at story time, the dances as a break during the day, and the songs during sing along time. The review is typically done on Fridays. It is intended to supplement and not substitute for the normal classroom literacy curriculum.

Prior to this evaluation, the BWC had been implemented in several schools and had received positive feedback from teachers who used it. At the time of the evaluation, the cost of the Big Word Club to schools was $60.00 per classroom regardless of the size of the class. As of 2023, the program offers an online subscription option (*BWC-Plus*) that costs $10/month.

3 Experimental Design

3.1 Sample Recruitment

Preschool and kindergarten teachers from the United States were recruited online via Facebook to apply to have their school participate in the study. In total, 637 teachers applied, of whom 260 were from schools that clustered in Arizona, Colorado, and Texas, while the rest were scattered throughout the United States. Because we had to collaborate with 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 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: Denver, Phoenix, Houston, and San Antonio. Finally, we eliminated home-based childcare centers. This left 96 eligible schools in which at least one teacher had volunteered.

We then emailed the principals of these 96 schools to tell them about the evaluation and to ask if they agreed that teachers could participate. From these 96 emails, 53 principals agreed that their school could participate in the evaluation. All teachers at the school, regardless of whether they had indicated interest on Facebook, were invited to participate. The main reason principals gave for not participating is that the school district rules prevent participation in outside research projects or that school district rules require a lengthy application process for outside researchers.

3.2 Randomization to Treatment

We randomized these 53 schools to either the treatment (26) or the control group (27). We had no contact with control schools other than assessments. Treatment school teachers were

all given free access to the BWC program and encouraged to use it as intended (one word per day). Every treatment teacher logged into the BWC platform at least once. However, the platform could not track video viewing. So, it might be the case that some teachers did not fully utilize the platform. On the other hand, a teacher might have logged into the platform just once but used it regularly without ever logging out. Because we do not have reliable data on compliance, we estimate an Intent-to-Treat effect rather than a Treatment-on-Treated effect. Arguably, the ITT effect is more policy-relevant, given that teachers may not fully comply in practice even when they are given free access to the platform.

After randomization, six of the 53 schools withdrew from the study just before the treatment began. One of these schools was in the treatment group and five were in the control group. The two most common reasons that schools gave for withdrawing were scheduling conflicts with the assessment period and unwillingness to allow external personnel to assess students. Because we did not re-randomize, the final sample was 47 schools – 21 in the control group and 26 in the treatment group. In section 4.1, we see that the treatment and control groups are still balanced across all available covariates, which gives us confidence that there are no internal validity concerns. This distribution of the 818 students in these 47 schools by state is shown in Appendix Table A1.

(Click to view: Appendix Table A1).

3.3 Assessments

Our main outcome is an assessment we created (which we call the "BWC Assessment") that is a vocabulary test of words covered in the first 4 months (16 weeks) of the BWC program. The treatment began in November 2017, and we gave the BWC assessment twice: in March 2018

(hereafter, *Post-Test*) and May 2018 (hereafter, *Follow-up*). These correspond to four and six months after the treatment began respectively. No baseline assessment was conducted.

Our secondary outcome is the Peabody Picture Vocabulary Test (hereafter, *PPVT*). The PPVT is a widely-used test that evaluates receptive vocabulary (Dunn & Dunn, 2007). This assessment was given in May 2018 along with the *Follow-up*, which was six months after the treatment began. The purpose of this assessment was to see whether exposure to the BWC program had an impact on vocabulary skills more generally. On one hand, the program may improve general vocabulary skills by increasing student interest in learning words. On the other hand, we may see a decrease in general vocabulary skills if exposure to the unique words in the BWC came at the cost of spending time learning more commonly used words.

We developed the BWC Assessment to be similar in presentation to the PPVT. For each word, a page with 4 pictures was shown to a child. One of the pictures depicted the target word and the three other 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 is the total number of correctly selected words. The BWC protocol for assessment administration was identical to the PPVT protocol in terms of the prompts given to students. For instance, assessors were instructed to use neutral language like "okay" and "thank you" rather than "good job" or "well done."

One difference between the PPVT and the BWC assessment is that the words in the PPVT are presented in increasing order of difficulty, whereas the words in the BWC assessment are not ordered by difficulty. Another difference was that the BWC assessment was administered as a test booklet with paper score sheets for assessors, while the PPVT was administered on iPads through Pearson's Q-Interactive digital platform.

In developing the BWC assessment, pilot tests showed that children were able to attempt 38 words in 5 minutes, on average. The assessment covered words from the first 16 weeks of the BWC program (80 possible words). After omitting 20 words that were either difficult to depict with a static image (e.g., silent) or had multiple meanings (e.g., "healthy" can be interpreted as "in good health" or "good for you"), we randomly selected 38 out of the remaining 60 words to be used in our assessment. The assessment is available in the online appendix.

The BWC program includes words of varying familiarity and difficulty for preschool and kindergarten children. We use two approaches to characterize the difficulty level of words on the BWC assessment. One approach was to use the Brigham Young University (BYU) iWeb corpus database, which lists the frequency of words appearing in over 22 million web pages. Because frequently used words are more likely to be known by children, frequency in this database can be considered a measure of difficulty level. The second approach was to use the varying "difficulty tiers" of words, as outlined in the Common Core State Standards. Tier 1 words are defined as *commonly used, basic* vocabulary; Tier 2 words are *high-utility academic* vocabulary; and Tier 3 words are *domain-specific academic* vocabulary.

Appendix Table A2 shows the BYU iWeb Corpus frequency as well as the Common Core State Standards Tier level for the 38 words in the BWC assessment. The words are shown in order of their appearance on the assessment. The assessment contains words with a wide range of frequency levels. Approximately half of the words are Tier 3, implying that the words assessed are relatively difficult. This is consistent with the aim of the Big Word Club program to teach "big" words to children.

(Click to view: Appendix Table A2).

3.4 Covariates

Three demographic variables were available for each student from administrative data: *Female, Kindergarten,* and *Age*. *Female* is a binary variable with a value of 1 if the child is female; *Kindergarten* is binary, with a value of 1 if the child is in kindergarten and 0 if the child is in preschool; *Age* is the child's age in years as of May 2018.

Three additional school-level variables are also available: *Free Lunch* proportion, *Title 1* Status, and *Private School* status. *Free Lunch* is the proportion of students in that child's school that receive free or reduced-price lunch; *Title I* and *Private School* are binary with a value of 1 if the child's school is Title-I or Private respectively.

We use all six of these available variables as covariates to improve the precision of our estimates. The pre-registration for this RCT did not include a pre-analysis plan, and we show the results with and without controlling for these variables.

3.5 Sample Attrition

Given that testing was conducted one-on-one with individual assessors, it was cost-prohibitive for us to test every student in a school. Given that the randomization was at the school level, there isn't a substantial power gain from a small increase in the number of students within each cluster. Based on power calculations, our target was to test between 15-20 students per school. During the *Post-Test* assessment in March 2018, we tested an average of 17 students per school, giving us a total sample size of 818 students.

When we administered the *Follow-up* and *PPVT* tests two months later in May 2018, we were able to test 609 of the students we had tested in March. Attrition was primarily because of the high number of absences as the school year was close to ending and families left for summer vacation early. Additionally, some children who were tested in March were unable to stay

focused enough to take the assessment in May. There was no significant difference in the proportion of treatment versus control students who did not take the *Follow-Up* or *PPVT* tests. Because attrition is primarily based on student absence, this is unlikely to be systematically different based on treatment status; the family's decision for a student to be absent is unlikely to be related to their teacher being assigned to play a few minutes of videos. This gives us confidence in the internal validity of the results.

Another issue was fatigue, which led some students to only be able to take one of the two assessments – the *Follow-up* and *PPVT* – in May. Of the 609 students who were tested, 603 took the *Follow-up*, 591 took the *PPVT*, and 585 took both assessments. In estimating the treatment effect on the *Follow-up* and *PPVT*, we focus our analysis on the 585 students who took both tests so that we are looking at the same group of children across regressions. That said, the statistical and economic significance of the regression coefficients are similar whether we look only at the 585 students who took both assessments, or the 603 or 591 students respectively who took just one of the two assessments.

4 Results

4.1 Descriptive Results

Table 1 shows the descriptive statistics and balance tests for the six available covariates. There is no significant difference between treatment and control groups across all covariates. Overall, approximately half of the children in the sample are female, and most students attend Title-I schools. The average student is between 5 and 6 years old at the time of the *Follow-up*, and approximately half are in kindergarten.

(Table 1 here)

The *No Follow-up* variable in the last row is the attrition rate; it is binary with a value of 1 if the child took the *Post-Test* assessment but did not take both the *Follow-up* and *PPVT* assessments. We see that the control and treatment groups had attrition rates of 29% and 28% respectively, with no statistically significant difference. This balance mitigates internal validity concerns when analyzing the results of the *Follow-up* and *PPVT* assessments.

4.2 Post-Treatment Regression Results

Four months after the treatment began, 818 students were given the BWC assessment on words covered in the first four months of the BWC program. Out of 38 words, the control group students correctly identified an average of 22.7 words while the treatment group identified 24.3. The standard deviation was around 5.5 words for both groups. We standardize the outcome by subtracting the control group's mean from each student's score and then dividing by the standard deviation. We estimate the following regression:

$$Y_i = \alpha + \beta T_i + \Gamma X_i + S_j + \varepsilon_i$$

Where Y_i is the standardized *Post-Test* score for student i, T_i is the student's treatment status, X_i is the vector of available covariates, S_j is the state fixed effect for state j, and ε_i is the standard error clustered at the school-level. The results are shown in Table 2.

(Table 2 here)

The first column in Table 2 shows that the treatment students scored 0.28 SD more than control students, but that this difference is not statistically significant. The second column estimates the treatment effect with state fixed effects included, and the third column also

includes the six available covariates. Including these controls does not change the estimated treatment effect, but substantially reduces the standard error of the estimate. In this specification, the treatment effect is 0.3 SD and is significant at the $\alpha = .01$ level. The covariate coefficients in the third column show there is no significant difference in performance by gender, but that older students and students in high-SES schools perform better.

4.3 Follow-up Regression Results

As described in Section 3.5, only 585 out of the 818 children who took the *Post-Test* were able to take the *Follow-up* and *PPVT* tests two months later (hereafter, *follow-up sample*). The descriptive statistics for the follow-up sample are like those of the full sample: about half of the children are female, about half are in kindergarten, and a majority are in Title-I schools. Appendix Table A3 shows that all available covariates are balanced. Additionally, the treatment effect on the *Post-Test* score for the follow-up sample of 585 is 0.313 SD, and significant at the $\alpha = .01$ level, as shown in appendix Table A4. This is like the 0.307 SD treatment effect of the full sample shown in Table 2.

(Click to view: Appendix Table A3)

(Click to view: Appendix Table A4)

The *Follow-up* assessment administered in May 2018 was identical to *Post-Test* from two months earlier. We estimate the following regression:

$$Y_i = \alpha + \beta T_i + \Gamma X_i + S_j + \varepsilon_i$$

Where Y_i is the standardized *Follow-up* score for student i, T_i is the student's treatment status, X_i is the vector of available covariates, S_j is the state fixed effect for state j, and ε_i is the standard error clustered at the school-level. The results are shown in Table 3.

(Table 3 here)

The first column in Table 3 shows that the treatment students scored 0.25 SD more than control students on the *Follow-up*, but that this difference is not statistically significant. The second and third columns include state fixed effects and six available covariates, respectively. As was the case with the *Post-Test* regression in Table 2, including these controls does not change the estimated treatment effect, but improves precision. In this specification, the treatment effect is 0.27 SD and is significant at the $\alpha = .05$ level. This indicates that the treatment effect of the first 4 months of the BWC program was mostly retained two months later.

4.4 PPVT Regression Results

In addition to the *Follow-up* BWC assessment in May 2018, we also administered the *PPVT* – a widely used vocabulary test. The scale of the PPVT is from 20 to 160 points. The control students scored an average of 88.7 points, while the treatment students scored an average of 90.7 points. The standard deviation was approximately 25.5 points. We standardize the outcome by subtracting the control group's mean from each student's score and then dividing by the standard deviation. We estimate the following regression:

$$Y_i = \alpha + \beta T_i + \Gamma X_i + S_j + \varepsilon_i$$

Where Y_i is the standardized *PPVT* score for student i, T_i is the student's treatment status, X_i is the vector of available covariates, S_j is the state fixed effect for state j, and ε_i is the standard error clustered at the school-level. The results are shown in Table 4.

(Table 4 here)

The first column in Table 4 shows that the treatment students scored 0.08 SD more than control students on the *PPVT*, but that this difference is not statistically significant. The second and third columns include state fixed effects and six available covariates, respectively. While the inclusion of these controls improves precision, the treatment effect is not statistically significant. However, we cannot rule out a large positive impact of the BWC program on PPVT scores. This implies that there is not enough precision in these data to make confident claims about the program's impact on general vocabulary knowledge.

4.5 Exploratory Heterogeneity Analysis

The study design does not have enough statistical power to conduct heterogeneity analysis due to multiple hypothesis testing concerns (List et al., 2019). With that in mind, an exploratory analysis of the differences in treatment effects by gender, grade-level, and SES are shown in tables 5, 6, and 7 respectively. These tables estimate the following regression:

$$Y_i = \alpha + \beta_1 T_i + \beta_2\, T_i X_i + \Gamma X_i + S_j + \varepsilon_i$$

Where Y_i is either the standardized *Post-Test*, *Follow-up*, or *PPVT* score for student i, T_i is the student's treatment status, X_i is the vector of available covariates, S_j is the state fixed effect

for state j, and ε_i is the standard error clustered at the school-level. The estimates of β_2 will indicate whether there is a difference in treatment effect by covariate X.

(Table 5 here)

Table 5 shows that there was no significant difference in the *Post-Test* treatment effect for boys and girls. However, girls had a significantly higher treatment effect on the *Follow-up* and *PPVT* assessments. While this may indicate that girls are more likely to retain words learned in the program and improve general vocabulary skills, the difference in treatment effect could also be due to spurious low performance by girls in the control group during the May 2018 test day. This is indicated by the negative coefficient on the *Female* variable for the *Follow-up* and *PPVT* assessments, significant at the $\alpha = .10$ level. Figure 1 shows the standardized scores for boys and girls separately on the *Post-Test* and *Follow-up*.

(Figure 1 here)

The point estimates in Table 6 and 7 show that the treatment effect was higher for older students and for students in schools with a low proportion of students on free or reduced-price lunch. However, these differences in treatment effect were not statistically significant.

(Table 6 here)

(Table 7 here)

5 Discussion

The results from this study indicate that consistent, light-touch educational media employed in classrooms can produce a modest amount of sustained learning. While the program

does not substitute time away from other non-targeted vocabulary words, our data lacks the precision necessary to determine whether the program improves vocabulary skills more generally, as measured by the PPVT.

One limitation with this study is that a reliable compliance measure is not available which prohibits us from computing a Treatment-on-Treated (TOT) estimate of the program. However, the Intent-to-Treat (ITT) estimates in this study are more representative of treatment effects that one would find at scale where compliance is imperfect. Another limitation is that no baseline test was conducted, which reduces the precision of our estimates. However, the fact that treatment is randomly assigned and that baseline covariates are balanced gives us confidence in the internal validity of the results.

These results could help schools decide whether to implement these types of light-touch educational media programs at scale (Mayer et al., 2021; Toma & Bell, 2022). Programs that are highly time-intensive may have higher impacts, but also have a higher opportunity cost that should be considered when deciding between alternative uses of that time. For instance, a program that is only 5 minutes per day might not change how a teacher spends the rest of his or her class time, which is less likely to be the case for a program that is 30-60 minutes per day.

Moreover, this classroom intervention could be easily delivered to children at home and could yield even bigger benefits. Teachers or school could make the videos available to parents on a website or by texting them to parents' phones. This would create an additional channel through which schools engage and communicate with parents (Avvisati et al., 2014; Kalil, Liu, et al., 2023; Kalil, Mayer, et al., 2023; Shah et al., 2022). However, watching videos virtually may not be as effective as watching them in class (Kofoed et al., 2021; List & Shah, 2022). Future work could compare the impact of educational media when used by teachers in the classroom to

their impact when used by parents at home, or the joint impact of these complementary approaches.

References

Avvisati, F., Gurgand, M., Guyon, N., & Maurin, E. (2014). Getting parents involved: A field experiment in deprived schools. *Review of Economic Studies*, *81*(1), 57–83.

Banerjee, A., Barnhardt, S., & Duflo, E. (2015). Movies, margins, and marketing: Encouraging the adoption of iron-fortified salt. In *Insights in the Economics of Aging* (pp. 285–306). University of Chicago Press.

Banerjee, A., La Ferrara, E., & Orozco-Olvera, V. H. (2019). *The entertaining way to behavioral change: Fighting HIV with MTV*. National Bureau of Economic Research.

Barsoum, G., Crépon, B., Gardiner, D., Michel, B., & Parienté, W. (2022). Evaluating the Impact of Entrepreneurship Edutainment in Egypt: An Experimental Approach. *Economica*, *89*(353), 82–109.

Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal*, *107*(3), 251–271.

Berg, G., & Zia, B. (2017). Harnessing emotional connections to improve financial decisions: Evaluating the impact of financial education in mainstream media. *Journal of the European Economic Association*, *15*(5), 1025–1055.

Bogatz, G. A., & Ball, S. (1971). *The Second Year of Sesame Street: A Continuing Evaluation. Volume 1*. Educational Testing Service. https://eric.ed.gov/?id=ED122800

Coville, A., Di Maro, V., Dunsch, F., & Zottel, S. (2019). The Nollywood Nudge: An Entertaining Approach to Saving. *World Bank Policy Research Working Paper*, *8920*.

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.

Diaz-Guerrero, R., & Holtzman, W. H. (1974). Learning by televised Plaza Sesamo in Mexico. *Journal of Educational Psychology*, *66*(5), 632.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.

Hart, B., & Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Brookes.

Kalil, A., Liu, H., Mayer, S., Rury, D., & Shah, R. (2023). Nudging or Nagging? Conflicting Effects of Behavioral Tools. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, *2023–02*.

Kalil, A., Mayer, S., & Shah, R. (2023). Scarcity and Inattention. *Journal of Behavioral Economics for Policy*, *7*(1), 35–42.

Kearney, M. S., & Levine, P. B. (2015). Media influences on social outcomes: The impact of MTV's 16 and pregnant on teen childbearing. *American Economic Review*, *105*(12), 3597–3632.

Kearney, M. S., & Levine, P. B. (2019). Early childhood education by television: Lessons from Sesame Street. *American Economic Journal: Applied Economics*, *11*(1), 318–350.

Kofoed, M., Gebhart, L., Gilmore, D., & Moschitto, R. (2021). Zooming to class?: Experimental evidence on college students' online learning during Covid-19. *Online Learning During COVID-19. IZA Discussion Paper*, *14356*.

Lee, J. (2008). The educational impact of Sisimpur: Results of an experimental study of children's learning. *Annual Meeting of the International Communication Association, San Francisco, CA, October*, *23*.

List, J. A. (2022). *The voltage effect: How to make good ideas great and great ideas scale*. Currency.

List, J. A., & Shah, R. (2022). The impact of team incentives on performance in graduate school: Evidence from two pilot RCTs. *Economics Letters*, *221*, 110894.

List, J. A., Shaikh, A. M., & Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, *22*, 773–793.

Mares, M.-L., & Pan, Z. (2013). Effects of Sesame Street: A meta-analysis of children's learning in 15 countries. *Journal of Applied Developmental Psychology*, *34*(3), 140–151.

Mayer, S., Shah, R., & Kalil, A. (2021). How cognitive biases can undermine program scale-up decisions. In *The Scale-Up Effect in Early Childhood and Public Policy* (pp. 41–57). Routledge.

Moran, C., Hilgendorf, D., Zhao, V., Al-Ogaily, D., Yoshikawa, H., Schwartz, K., Rafla, J., Molano, A., Strouf, K., Khanji, M., Abu Seriah, R., Al Aabed, M., Fityan, R., Sloane, P., Hussein, L., Hidayah, D., Shukri, M., Sharawi, T., Foulds, K., … Behrman, J. (2023). *Impacts and Costs of a Mass Media Program on Children's Emotion Knowledge, Recognition, and Regulation in Jordan: A Cluster-Randomized Controlled Trial*.

Neuman, S. B., Wong, K. M., Flynn, R., & Kaefer, T. (2019). Learning vocabulary from educational media: The role of pedagogical supports for low-income preschoolers. *Journal of Educational Psychology*, *111*(1), 32.

Penuel, W. R., Bates, L., Gallagher, L. P., Pasnik, S., Llorente, C., Townsend, E., Hupert, N., Domínguez, X., & VanderBorght, M. (2012). Supplementing literacy instruction with a media-rich intervention: Results of a randomized controlled trial. *Early Childhood Research Quarterly*, *27*(1), 115–127.

Ravallion, M., Van De Walle, D., Dutta, P., & Murgai, R. (2015). Empowering poor people through public information? Lessons from a movie in rural India. *Journal of Public Economics*, *132*, 13–22.

Scarborough, H. S., Neuman, S., & Dickinson, D. (2009). Connecting early language and literacy to later reading (dis) abilities: Evidence, theory, and practice. *Approaching Difficulties in Literacy Development: Assessment, Pedagogy and Programmes*, *10*, 23–38.

Shah, R., Kalil, A., & Mayer, S. (2022). Engaging Parents with Preschools: Evidence from a Field Experiment. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, *2022–97*.

Stahl, S. A., & Nagy, W. E. (2007). *Teaching word meanings*. Routledge.

Suskind, D., Suskind, B., & Lewinter-Suskind, L. (2015). *Thirty million words: Building a child's brain*. Dutton.

Toma, M., & Bell, E. (2022). Understanding and Increasing Policymakers' Sensitivity to Program Impact. *Available at SSRN 4435532*.

Tables

Table 1: Covariate Summary Statistics and Balance Test (N=818)

	Control (N=364)	Treatment (N=454)	p-value
Age	5.57 (0.82)	5.71 (0.70)	0.484
Female	0.51 (0.50)	0.50 (0.50)	0.553
Kindergarten	0.42 (0.49)	0.48 (0.50)	0.698
Free Lunch	0.47 (0.35)	0.62 (0.34)	0.144
Private School	0.22 (0.41)	0.15 (0.35)	0.526
Title I School	0.58 (0.49)	0.70 (0.46)	0.401
No Follow-up	0.29 (0.45)	0.28 (0.45)	0.836

Note. Standard deviations are in parentheses below variable means. The p-value column shows the p-value of the coefficient of a regression of each covariate on treatment status, with standard errors clustered at the school level. Age is the child's age in years as of May 2018. Female is a binary variable with a value of 1 if the child is female. Kindergarten is a binary variable with a value of 1 if the child is in kindergarten and 0 if the child is in preschool. "Free Lunch" is the proportion of students in that child's school that receive free or reduced-price lunch. Private school and Title I school are binary variables with a value of 1 if the child's school is private or Title-I respectively. No Follow-up is a binary variable with a value of 1 if the child took the Post-Test assessment but did not take both the Follow-up and PPVT assessments. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Click to go back

Table 2: Treatment Effect (SD units) of the BWC program on Vocab Skills at Post-Treatment

	(1) Post-Test	(2) Post-Test	(3) Post-Test
Treat	0.283 (0.194)	0.264 (0.164)	0.307*** (0.0824)
Age			0.438*** (0.121)
Female			0.0501 (0.0541)
Kindergarten			0.490*** (0.170)
Free Lunch			-0.785** (0.310)
Private School			-0.148 (0.215)
Title I School			-0.0292 (0.194)
Observations	818	818	818
State FE	No	Yes	Yes
Covariates	No	No	Yes

Note. Standard errors (clustered at the school level) are in parentheses. The outcome is the student's score, in standard deviation units, on an assessment of words taught in the first 4 months of the BWC program, 4 months after the treatment began. Age is the child's age in years as of May 2018. Female is a binary variable with a value of 1 if the child is female. Kindergarten is binary, with a value of 1 if the child is in kindergarten and 0 if the child is in preschool. "Free Lunch" is the proportion of students in that child's school that receive free or reduced-price lunch. Private school and Title I school are binary with a value of 1 if the child's school is private or Title-I respectively. *** p<0.01, ** p<0.05, * p<0.10.

Click to go back

Table 3: Treatment Effect (SD units) of the BWC program on Vocab Skills at Follow-up

	(1) Follow-up	(2) Follow-up	(3) Follow-up
Treat	0.252 (0.206)	0.254 (0.177)	0.274** (0.105)
Age			0.516*** (0.0833)
Female			-0.0408 (0.0623)
Kindergarten			0.380** (0.150)
Free Lunch			-0.677** (0.304)
Private School			-0.324 (0.246)
Title I School			-0.246 (0.242)
Observations	585	585	585
State FE	No	Yes	Yes
Covariates	No	No	Yes

Note. Standard errors (clustered at the school level) are in parentheses. The outcome is the student's score, in standard deviation units, on an assessment of words taught in the first 4 months of the BWC program, 6 months after the treatment began. Age is the child's age in years as of May 2018. Female is a binary variable with a value of 1 if the child is female. Kindergarten is binary, with a value of 1 if the child is in kindergarten and 0 if the child is in preschool. "Free Lunch" is the proportion of students in that child's school that receive free or reduced-price lunch. Private school and Title I school are binary with a value of 1 if the child's school is private or Title-I respectively. *** p<0.01, ** p<0.05, * p<0.10.

Click to go back

Table 4: Treatment Effect (SD units) of the BWC program on PPVT Test at Post-Treatment

	(1) PPVT	(2) PPVT	(3) PPVT
Treat	0.0803	0.0826	0.112
	(0.210)	(0.175)	(0.0844)
Age			0.607***
			(0.0814)
Female			0.0599
			(0.0540)
Kindergarten			0.309**
			(0.131)
Free Lunch			-0.777**
			(0.295)
Private School			-0.199
			(0.177)
Title I School			-0.161
			(0.244)
Observations	585	585	585
State FE	No	Yes	Yes
Covariates	No	No	Yes

Note. Standard errors (clustered at the school level) are in parentheses. The outcome is the student's score on the PPVT-4 in standard deviation units, 6 months after the treatment began. Age is the child's age in years as of May 2018. Female is a binary variable with a value of 1 if the child is female. Kindergarten is binary, with a value of 1 if the child is in kindergarten and 0 if the child is in preschool. "Free Lunch" is the proportion of students in that child's school that receive free or reduced-price lunch. Private school and Title I school are binary with a value of 1 if the child's school is private or Title-I respectively. *** p<0.01, ** p<0.05, * p<0.10.

Click to go back

Table 5: Heterogeneity by Gender in Treatment Effect of the BWC program on Vocab Skills

	(1) Post-Test	(2) Follow-up	(3) PPVT
Treat	0.291**	0.146	-0.0801
	(0.115)	(0.123)	(0.103)
Female	-0.0330	-0.172*	-0.137*
	(0.0641)	(0.0966)	(0.0730)
Treat X Female	0.0395	0.236*	0.354***
	(0.100)	(0.125)	(0.0982)
Kindergarten	0.273*	0.374**	0.300**
	(0.138)	(0.151)	(0.133)
Age	0.518***	0.521***	0.614***
	(0.0816)	(0.0836)	(0.0822)
Free Lunch	-0.852**	-0.655**	-0.743**
	(0.317)	(0.304)	(0.294)
Private School	-0.311	-0.337	-0.218
	(0.197)	(0.245)	(0.178)
Title I School	-0.0368	-0.264	-0.189
	(0.248)	(0.243)	(0.243)
Observations	585	585	585
State FE	Yes	Yes	Yes
Covariates	Yes	Yes	Yes

Note. Standard errors (clustered at the school level) are in parentheses. The outcome in the first column is the student's score, in standard deviation units, on an assessment of words taught in the first 4 months of the BWC program, 4 months after the treatment began. The outcome in the second column is the student's score (in SD units) of this same test, 6 months after the treatment began. The outcome in the third column is the student's score (in SD units) on the PPVT-4, 6 months after the treatment began. Female is a binary variable with a value of 1 if the child is female. Kindergarten is binary, with a value of 1 if the child is in kindergarten and 0 if the child is in preschool. Age is the child's age in years as of May 2018. "Free Lunch" is the proportion of students in that child's school that receive free or reduced-price lunch. Private school and Title I school are binary with a value of 1 if the child's school is private or Title-I respectively. *** p<0.01, ** p<0.05, * p<0.10.

Click to go back

Table 6: Heterogeneity by Grade in Treatment Effect of the BWC program on Vocab Skills

	(1) Post-Test	(2) Follow-up	(3) PPVT
Treat	0.180	0.134	0.0496
	(0.137)	(0.188)	(0.151)
Kindergarten	0.0912	0.187	0.224
	(0.163)	(0.185)	(0.170)
Treat X Kinder	0.282	0.297	0.131
	(0.196)	(0.230)	(0.190)
Female	-0.0115	-0.0414	0.0596
	(0.0479)	(0.0616)	(0.0535)
Age	0.535***	0.535***	0.615***
	(0.0845)	(0.0867)	(0.0837)
Free Lunch	-0.762**	-0.578*	-0.733**
	(0.324)	(0.323)	(0.317)
Private School	-0.293	-0.308	-0.192
	(0.195)	(0.249)	(0.177)
Title I School	-0.102	-0.317	-0.193
	(0.253)	(0.263)	(0.262)
Observations	585	585	585
State FE	Yes	Yes	Yes
Covariates	Yes	Yes	Yes

Note. Standard errors (clustered at the school level) are in parentheses. The outcome in the first column is the student's score, in standard deviation units, on an assessment of words taught in the first 4 months of the BWC program, 4 months after the treatment began. The outcome in the second column is the student's score (in SD units) of this same test, 6 months after the treatment began. The outcome in the third column is the student's score (in SD units) on the PPVT-4, 6 months after the treatment began. Kindergarten is binary, with a value of 1 if the child is in kindergarten and 0 if the child is in preschool. Female is a binary variable with a value of 1 if the child is female. Age is the child's age in years as of May 2018. "Free Lunch" is the proportion of students in that child's school that receive free or reduced-price lunch. Private school and Title I school are binary with a value of 1 if the child's school is private or Title-I respectively. *** p<0.01, ** p<0.05, * p<0.10.

Click to go back

Table 7: Heterogeneity by SES in Treatment Effect of the BWC program on Vocab Skills

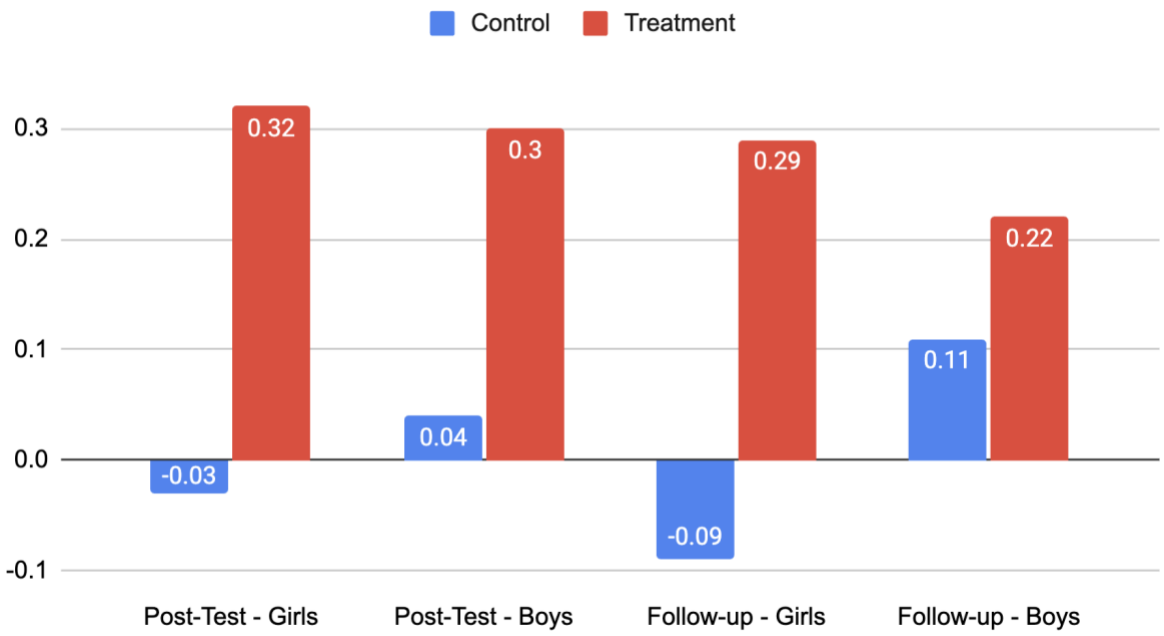
	(1) Post-Test	(2) Follow-up	(3) PPVT
Treat	0.430***	0.476**	0.211
	(0.151)	(0.208)	(0.154)
Free Lunch	-0.702*	-0.411	-0.645**
	(0.370)	(0.378)	(0.321)
Treat X Free Lunch	-0.216	-0.373	-0.184
	(0.276)	(0.355)	(0.261)
Female	-0.0148	-0.0473	0.0567
	(0.0465)	(0.0602)	(0.0533)
Age	0.508***	0.501***	0.599***
	(0.0792)	(0.0799)	(0.0807)
Kindergarten	0.279**	0.388**	0.313**
	(0.137)	(0.148)	(0.132)
Private School	-0.312*	-0.330	-0.202
	(0.185)	(0.222)	(0.171)
Title I School	-0.0657	-0.301	-0.189
	(0.257)	(0.254)	(0.246)
Observations	585	585	585
State FE	Yes	Yes	Yes
Covariates	Yes	Yes	Yes

Note. Standard errors (clustered at the school level) are in parentheses. The outcome in the first column is the student's score, in standard deviation units, on an assessment of words taught in the first 4 months of the BWC program, 4 months after the treatment began. The outcome in the second column is the student's score (in SD units) of this same test, 6 months after the treatment began. The outcome in the third column is the student's score (in SD units) on the PPVT-4, 6 months after the treatment began. "Free Lunch" is the proportion of students in that child's school that receive free or reduced-price lunch. Female is a binary variable with a value of 1 if the child is female. Kindergarten is binary, with a value of 1 if the child is in kindergarten and 0 if the child is in preschool. Age is the child's age in years as of May 2018. Private school and Title I school are binary with a value of 1 if the child's school is private or Title-I respectively. *** p<0.01, ** p<0.05, * p<0.10.

Click to go back

Figures

Figure 1: Standardized Post-Test and Follow-up scores for boys and girls separately (N=585)



Note. Standard errors (clustered at the school level) are in parentheses. The outcome in the first column is the student's score, in standard deviation units, on an assessment of words taught in the first 4 months of the BWC program, 4 months

Click to go back

Online Appendix

An online appendix that contains the dataset, codebook, and assessments is available here: https://sites.google.com/view/bip-bwc-online-appendix/home

Appendix Tables

Table A1: Distribution of schools and students by state

State	Number of Schools	Number of Students
Arizona	10	181
Colorado	20	359
Texas	17	278
Total	**47**	**818**

Click to go back

Table A2: Frequency of Words and Tier for words in the BWC Assessment

	BWC Word	Frequency	Tier
1	Fossil	5508	3
2	Stomp	798	2
3	Vegetable	9855	1
4	Wings	13584	1
5	Rainforest	1214	3
6	Equator	1304	3
7	Primate	842	3
8	Slither	332	2
9	Gargantuan	675	3
10	Unique	31562	3
11	Nature	79112	2
12	Webbed	352	2
13	Fun	54375	1
14	Herd	4657	2
15	Front	133697	1
16	Alive	34360	1
17	Ocean	25220	1
18	Together	171628	1
19	Camouflage	1920	3
20	Wiggle	1202	1
21	Exercise	37728	1
22	Umbrella	4581	1
23	Plankton	538	3
24	Stegosaurus	67	3
25	Photosynthesis	815	3
26	Tuber	242	3
27	Tilt	3112	2
28	Complete	51811	3
29	Academia	1801	3
30	Bramble	274	3
31	Creature	10390	2
32	Scaly	446	3
33	Bat	8592	1
34	Prehensile	104	3
35	Slip	11604	2
36	Bask	550	3
37	Surf	4461	2
38	Symbiotic	717	3

Note. Frequency denotes the frequency of each word's occurrence in the BYU iWeb corpus, and Tier refers to the difficulty tier of each word based on the tier description in the Common Core State Standards.

Click to go back

Table A3: Covariate Summary Statistics and Balance Test (N=585)

	Control (N=259)	Treatment (N=326)	p-value
Age	5.56 (0.80)	5.69 (0.67)	0.507
Female	0.54 (0.50)	0.52 (0.50)	0.466
Kindergarten	0.43 (0.50)	0.47 (0.50)	0.798
Free Lunch	0.48 (0.35)	0.63 (0.35)	0.159
Private School	0.22 (0.41)	0.15 (0.36)	0.564
Title I School	0.59 (0.49)	0.69 (0.46)	0.491

Note. Standard deviations are in parentheses below variable means. The p-value column shows the p-value of the coefficient of a regression of each covariate on treatment status, with standard errors clustered at the school level. Age is the child's age in years as of May 2018. Female is a binary variable with a value of 1 if the child is female. Kindergarten is a binary variable with a value of 1 if the child is in kindergarten and 0 if the child is in preschool. "Free Lunch" is the proportion of students in that child's school that receive free or reduced-price lunch. Private school and Title I school are binary variables with a value of 1 if the child's school is private or Title-I respectively. *** p<0.01, ** p<0.05, * p<0.10.

Click to go back

Table A4: Follow-up Sample Treatment Effect (SD units) of the BWC program on Vocab Skills

	(1) Post-Test	(2) Post-Test	(3) Post-Test
Treat	0.276	0.281*	0.313***
	(0.184)	(0.159)	(0.0872)
Age			0.517***
			(0.0816)
Female			-0.0110
			(0.0487)
Kindergarten			0.274*
			(0.138)
Free Lunch			-0.856***
			(0.317)
Private School			-0.309
			(0.196)
Title I School			-0.0337
			(0.245)
Observations	585	585	585
State FE	No	Yes	Yes
Covariates	No	No	Yes

Note. Standard errors (clustered at the school level) are in parentheses. This table shows the results of the same regressions as those of Table 2, but restricting the sample to only those students who also took the follow-up test. The outcome is the student's score, in standard deviation units, on an assessment of words taught in the first 4 months of the BWC program, 4 months after the treatment began. Age is the child's age in years as of May 2018. Female is a binary variable with a value of 1 if the child is female. Kindergarten is binary, with a value of 1 if the child is in kindergarten and 0 if the child is in preschool. "Free Lunch" is the proportion of students in that child's school that receive free or reduced-price lunch. Private school and Title I school are binary with a value of 1 if the child's school is private or Title-I respectively. *** p<0.01, ** p<0.05, * p<0.10.

Click to go back