Technical Report # 1902

ReadWorks Article-A-Day™:

Using a Maze Assessment to Test the Impact

Of Building Background Knowledge on

Reading Comprehension

Susanne Nobles – ReadWorks

Daniel Anderson - BRT

Manjula Raman - ReadWorks

Katy Laird – ReadWorks

Gerald Tindal - BRT

University of Oregon

ReadWorks®

Published by

Behavioral Research and Teaching
University of Oregon • 175 Education
5262 University of Oregon • Eugene, OR 97403-5262
Phone: 541-346-3535 • Fax: 541-346-5689
http://brt.uoregon.edu

Abstract

This study was premised on the importance of vocabulary in comprehending text. Critical findings from both the National Reading Plan (NRP) and the National Reading Technical Assistance Center (NRTAC) frame this study, both in the intervention that was implemented and in the manner in which outcomes were measured. Using expository passages developed by ReadWorks, teachers implemented an 'Article-A-Day' with students reading brief expository passages. In this particular study, the passages focused on endangered plants and their survival, though the full domain of passages available in ReadWorks is extensive and addresses many other topics. The primary question was whether this strategy was more effective in supporting reading comprehension than when students did not consistently read passages on a daily basis. For five days, students read successive passages and, within two days of reading the last passage, completed an assessment that was based on a similar (but unfamiliar) passage with key targeted words omitted (using a maze format). Significant differences were found in the performances of these two groups, with students in the Article-A-Day treatment group performing higher than students in the control group. The most important implications of this study include the systematic focus on building background knowledge, including vocabulary, in reading comprehension instruction in which the active ingredients reflect findings from previous research.

Introduction

Nearly 20 years ago, the National Reading Panel (NICHD, 2000) outlined the primary areas to address in teaching children to read, using a five component framework: alphabetics (phonemic awareness and phonics), fluency, vocabulary, and text comprehension. In this study, we address vocabulary. As the subgroup from the NRP noted: "Vocabulary occupies an important position in learning to read. As a learner begins to read, reading vocabulary encountered in texts is mapped onto the oral vocabulary the learner brings to the task" (p. 4-15). This subgroup then cited a number of studies depicting comprehension gains and improvement on semantic tasks as the results of vocabulary learning, though the measurement of vocabulary can take many different forms. In citing important strategies for teaching vocabulary, they noted "that high frequency and multiple, repeated exposures to vocabulary material are important for learning gains. In accordance with this finding, a trend was also noted that extended and rich instruction of vocabulary (applying words to multiple contexts, etc.) was superior to less comprehensive methods" (p. 4-22). Other instructional practices that appeared influential included direct instruction, active engagement of students, pre-instruction, and use of rich contexts.

In a follow-up review of research on vocabulary instruction, the National Reading Technical Assistance Center (TRTAC, 2010) conducted a research synthesis of 14 studies. They cite three main conclusions:

1. "Higher frequency of exposure to targeted vocabulary words will increase the likelihood that young children will understand and remember the meanings of new words and use them more frequently" (p. 4).
2. "Explicit instruction of words and their meanings increases the likelihood that young children will understand and remember the meanings of new words" (p. 4).
3. "Questioning and language engagement enhance students' word knowledge" (p. 5).

With these two reviews, the study was framed on the passages in ReadWorks' research-based reading program Article-A-Day (AAD). AAD uses a 10-minute daily routine focused on three critical aspects of reading comprehension: increasing students' background knowledge, vocabulary, and reading stamina. AAD has been shown to help improve students' reading comprehension, and teachers have said that AAD is effective in improving students' background knowledge and vocabulary (Rockman, 2016, 2017).

Article-A-Day provides students with well-structured expository passages. The passages are relatively brief (250-350 words) so students can read them in a single session; they contain clear, direct sentences that provide content on a topic. The underlying assumptions behind the reading program include frequent (daily) reading of passages with a similar discourse structure having key vocabulary presented in context and active application of the information in a context.

Therefore, the goal of this pilot was to learn if AAD improved students' reading comprehension by giving them new background knowledge that they applied to understanding new texts with a similar structure. Specifically, our research question was whether a significant positive effect occurred from students completing AAD by reading a set of five topically connected articles as measured by maze vocabulary performance on a new text related to the same topic. To conduct a study of the effects from using AAD as an intervention, a dependent measure was designed to reflect a new passage that represented the same discourse structure with key vocabulary (on a new but related topic). This outcome assessment was developed using a passage focused on the Alula plant, which was a new article that students had not read previously but reflected the same text structure and similar topic (on endangered plants) as the previous AAD. The idea is that the vocabulary crosses among the articles and therefore students learn it more deeply. This type of study on expository text using a maze assessment has been

successfully deployed previously to reflect an emphasis on concepts and vocabulary present in subject areas (McCoy, Twyman, & Tindal, 2006; Twyman, McCleery, & Tindal, 2006).

Methods

In this section, we describe the participants, the treatment materials, the outcome measure, and the analyses that we conducted. The primary purpose of the study was to determine the (experimental) effects of using AAD, but we also were interested in creating and analyzing a measure for documenting these outcomes: a maze measure.

Participants

Volunteer teachers were recruited from the database of teachers registered on the ReadWorks digital website. The target goal of number of participants was 100 teachers in each of the control and treatment groups. This number was determined to be sufficiently large to draw initial conclusions from this pilot study.

First, the database of registered teachers was divided into two geographical areas: teachers in New York City (NYC) and teachers not in NYC but still in the U.S. The reason for this division was that a funder for the project, the Brooke Astor Fund for NYC Education, focuses specifically on supporting programs and activities that improve the quality of education in NYC. Therefore, the results of the experiment needed to be analyzed for students in NYC and students overall.

Within these two geographical areas, criteria were set to determine the teachers who would receive the first email request (sent between April 11-30, 2019) asking them to voluntarily participate in either the control or treatment group. A second round of email requests (sent May 14, 2019) was deemed necessary to recruit more participants. Specifically, the number of NYC participants who volunteered after the first request was much smaller than the target goal, and, in both NYC and the national geographic areas, the control group participation far exceeded the

treatment group participation. In the second round of requests, some of the criteria were eliminated to reach a larger number of possible participants. These are indicated below in parentheses.

Criteria for NYC Treatment Group: (a) NYC DOE school, (b) Grade 4 class, (c) digital class that had completed digital assignments (not a criterion for the second round), (d) four or more AAD uses, (e) no prior exposure to the Endangered Plants AAD set or any of the individual articles within the set (not a criterion for the second round).

Criteria for NYC Control Group: (a) NYC DOE school, (b) Grade 4 class, (b) digital class that had completed digital assignments (not a criterion for the second round), (c) no use (or single use) of an AAD set, (e) no exposure to the Endangered Plants AAD set or any of the individual articles within the set (not a criterion for the second round).

The same criteria were deployed for the national control and treament groups as was used with the NYC control and treatment groups with a few exceptions. The control group (a) was of sufficient size to not require a second round of requests, (b) had identified their school name in the 'school' field item within the ReadWorks, and (c) had not received outreach for a different ReadWorks experiment. Furthermore, the treatment group (a) had identified their school name in the 'school' field item within the ReadWorks, (b) had used AAD with 10 or more passages, and (c) had not received outreach for a different ReadWorks experiment.

Treatment Materials

The AAD program uses weekly sets of articles that are topically connected, such as a set of articles on storms or physical science. For this study, we used the AAD set about endangered plants because this is a topic that aligns with common curricular topics for the spring in fourth grade. We hoped this would make participating in the study less disruptive for teachers, thus increase the number who volunteered. While the AAD sets often have more than five articles to

allow for teacher and/or student choice, we limited the endangered plants set to only five articles for this pilot to ensure that students read similar content prior to taking the maze. The five Article-A-Day passages that were read by students in the treatment group included: (a) "The Most Stinky Flower on Earth," (b) "Survivor Trees," (c) "The Upside Down Tree of Life," (d) "Why Are Some Plants Endangered," and (e) "Steps You Can Take to Save Endangered Plants." Each of the passages was approximately 300 words in length and designed to be read in a single session. Teachers generally use the passages to supplement instruction.

Study Design

The study involved a control and treatment group. In the *control group*, teachers had their students complete the 15-question maze. Students had not read anything in preparation for the maze. Teachers were encouraged but not required to limit students to 10 minutes to take the assessment. In the *treatment group*, teachers were asked to implement the following protocol: (a) using the endangered plants AAD set, guide students through the AAD routine by reading one article a day for five days; (b) within 48 hours, ideally but no more than one week after this AAD routine was complete, have students complete the 15-question maze. Teachers were also encouraged, but not required, to limit students to 10 minutes to take the assessment. For the control group, we confirmed (within their use metrics on our digital platform) that they had not assigned the endangered plant AAD set, thus that they had followed the protocol. For the treatment group, we only included data in our analysis from teachers whose use metrics showed that they had both assigned the endangered plants AAD set and had their students complete the maze.

Outcome Measure

We developed a maze reading comprehension measure, a commonly used formative assessment to measure reading comprehension that has been established as both reliable and

valid for assessing reading comprehension progress (Fuchs & Fuchs, 2002; Marcotte & Hintze, 2009; Pierce, McMaster, & Deno, 2010; Shin, Deno, & Espin, 2000; Shin & McMaster, 2019). The maze generally uses a standardized cloze format in which students read a text with at least 300 words that, after the first sentence, has every seventh word removed and replaced with three choices: the correct word and two incorrect words that served as distractors (Fuchs & Fuchs, 2011, 1992). This format, however, has not always been strictly followed; for example, in some uses, the correct word and distractors are selected purposely to increase test difficulty (Shinn & Shinn, 2002).

Therefore, we adapted the maze format to try to test students' application of background knowledge and vocabulary to a new text. Instead of removing every seventh word, we removed keywords that would signal background knowledge about the AAD topic, as outlined in Liu, Sundstrom-Hebert, Ketterlin-Geller, and Tindal (2008) who used both classical and item response theory (IRT) to document reliability and validity data for this assessment type. Specifically, our maze reflected a vocabulary assessment that was contextual (Pearson, Hiebert, & Kamil, 2007); furthermore, the maze options (both correct and incorrect) were designed to be meaningful in sentences as both content-related and the same part of speech as the eliminated word (Parker, Hasbrouck, & Tindal, 1992). Finally, only three options were given, as per the meta-analysis conducted by Rodriguez (2005). In the construction of the options, we used Hiebert's (2005, 2006) semantic associations as a critical criterion in selecting vocabulary words for the maze (Pearson et al., 2007), though a second criterion was instructional presence using AAD within the teaching-learning cycle, given the treatment of previous AAD articles on endangered plants.

Our 15-item maze was about a specific endangered plant, the Alula, that had not been mentioned in any of the articles in the AAD endangered plants set. The passage was comparable in length and Lexile and formatted similarly to the articles in the endangered plants AAD set. The words we selected to omit in the maze were important words in the vocabulary of the study of endangered plants (e.g., damaged, invasive). In this way, the maze required students to use their background knowledge about endangered plants to choose the correct word as they read about an unfamiliar endangered plant.

Analyses

We have summarized participation data using traditional descriptive statistics to document participants and performance. Importantly, to ensure the treatment and control groups were implemented with fidelity, we analyzed two questions about their use of AAD and then removed those who confounded the treatment. Then we analyzed the two groups using multiple regression, eventually using three conditional models to control for time (both unconstrained and limited to one hour). Finally, because we were interested in analyzing the maze measure to determine its suitability for use in evaluating AAD, we conducted extensive analyses of the outcome measure, including an option analysis and its reliability.

Results

We summarize the results first with a preliminary exploration of the data set summarizing participation (of teachers and students) using descriptive statistics and graphic displays of student groups within teachers. We then summarize time spent completing the maze, again presenting graphs to show the distribution. Finally, we map teacher participation from the various sites in the country. The most important analysis is then presented in which we compare the treatment and control groups on the maze measure, first graphically displaying distributions and then analyzing statistical significance using multiple regression. Finally, we present an option analysis

of each item in which the proportion of students is displayed on the y-axis for each option with their total score displayed on the x-axis. The last analyses summarize the reliability of the maze.

Descriptive Statistics

The initial data analyses addressed the sample of teachers and students. Considerable variance was present in the number of students within each teacher's classroom. This distribution was bi-modal and generally reflected a group of teachers with under 10 students and another group with around 20-25 students. A few teachers had between 40 and 50 students. These distributions were generally similar for both Treatment (A) and Control (B) teachers.

Students from both the treatment and control groups took approximately the same amount of time (7 minutes); considerable variation, however, existed in the time taken to complete the maze (with a standard deviation of nearly 5 minutes).

The vast majority of teachers had not previously assigned the content or previously projected/printed the content. More Treatment than Control teachers had previously assigned the content; in contrast, more Control than Treatment teachers had previously projected or printed the content.

Table 1

Use of Article-A-Day by Intervention and Control Teachers

Exposure	Treatment	Control
Did not previously assign content	495	777
Previously assigned content	70	20
Exposure	Treatment	Control
Did not previously project or print content	559	753
Previously projected or printed content	6	44

In both cases of exposure, we removed any teachers from the comparative analysis if they had assigned, projected, or printed the content.

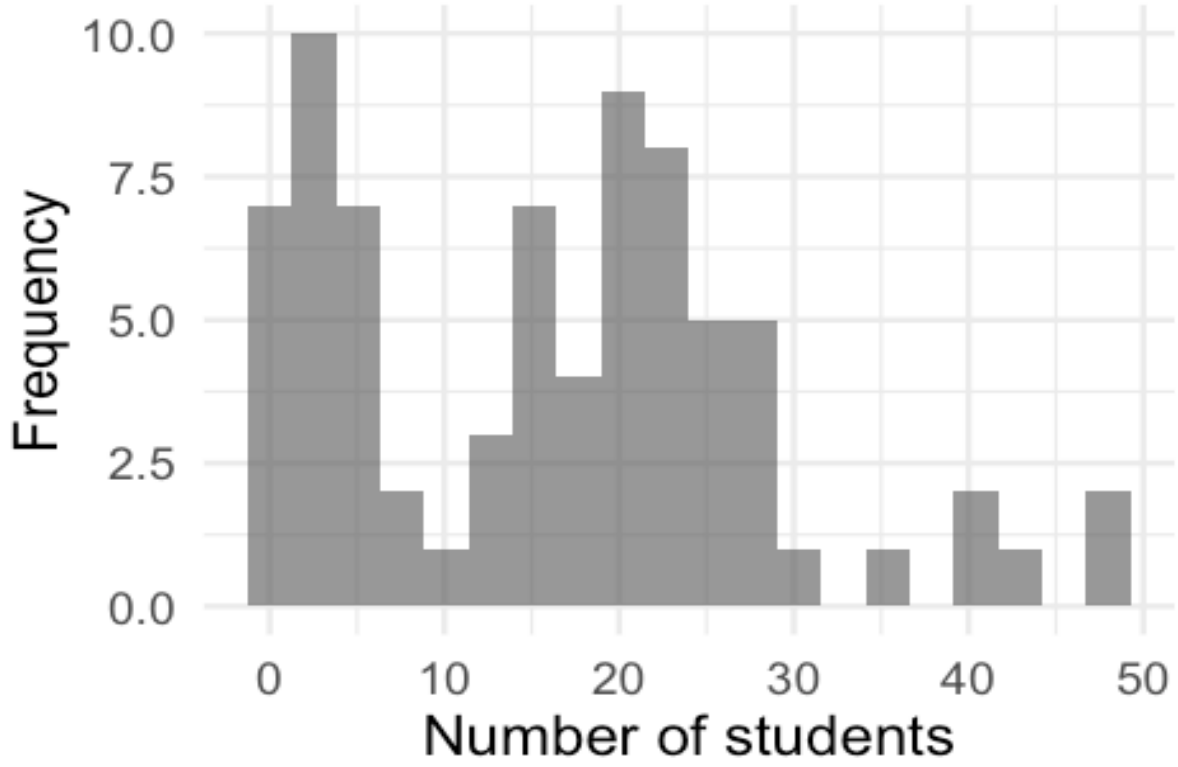

Figure 1. Distribution of number of students within teachers.

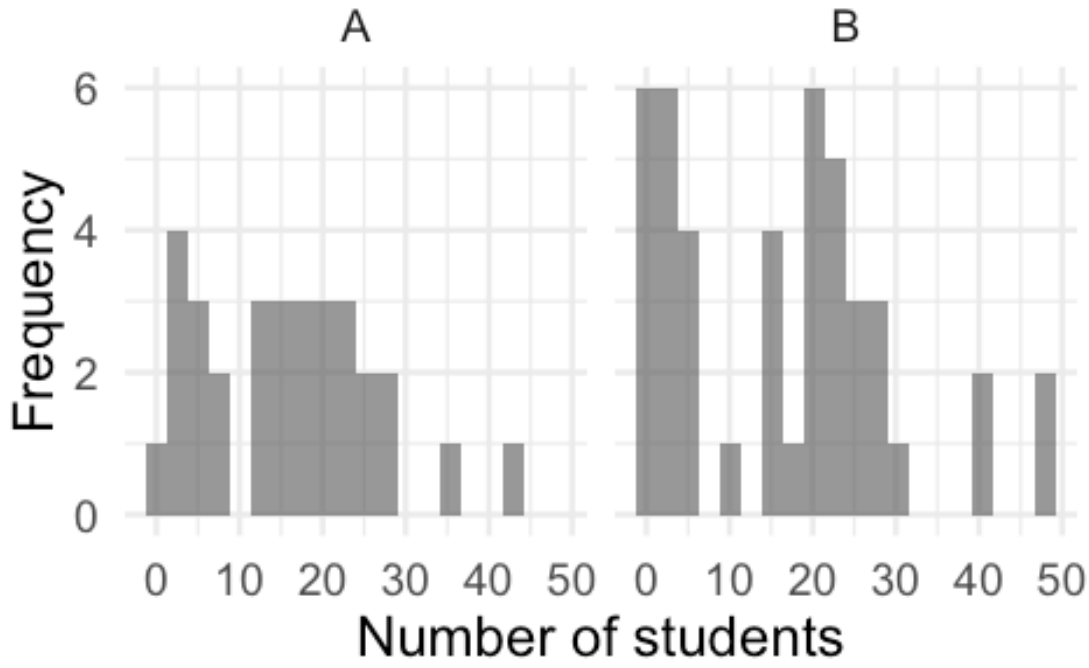

Figure 2. Distribution of number of students within teachers by treatment and control group.

Table 2

Time Taken (limited < 60 mins)

Group	Mean	SD
A	6.72 mins	4.44
B	6.89 mins	3.50

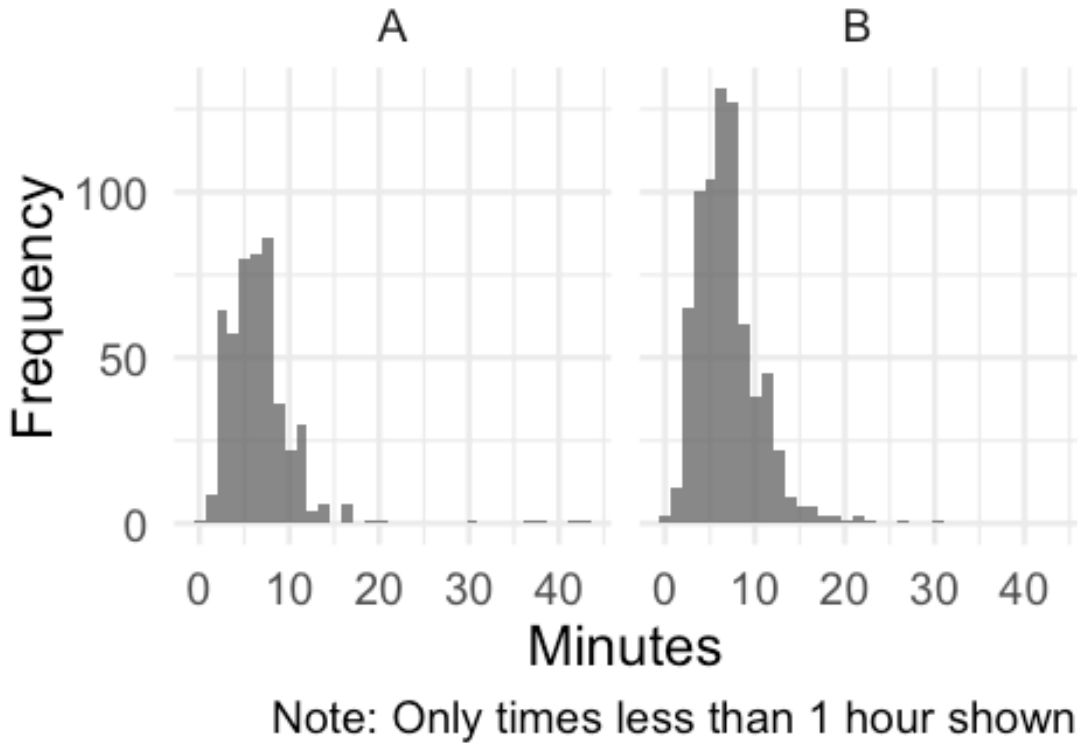

Figure 3. Amount of time taken to complete the maze.

Map of Participation

Using longitude and latitude information in the data file, a geo-spatial map was developed to show the sites participating in the study. The greatest concentration of sites were from the mid-west and east with California and Arizona representing the majority of the western sites (and single sites in Oregon, Washington, and Alaska).

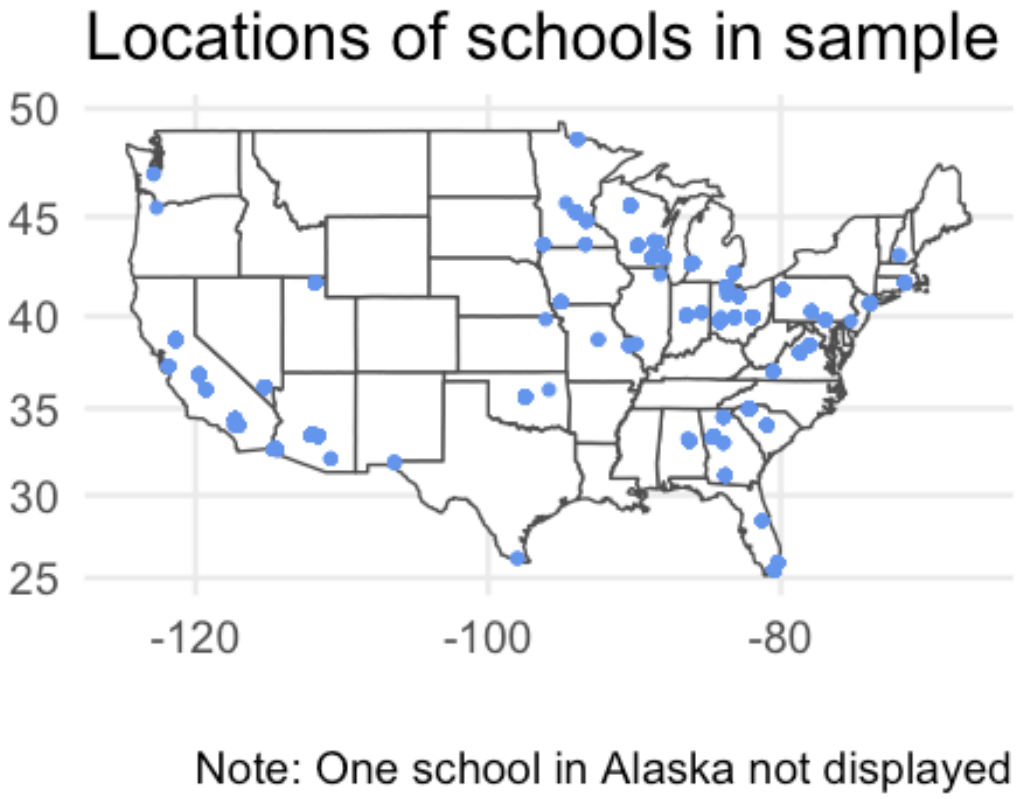

Figure 4. Geo-spatial distribution of participating sites.

Comparison of Treatment versus Control Groups

The distributions of performance for both the treatment and control groups overlapped considerably. When placed on the same plot, the Treatment distribution (green plot) was slightly higher than the Control distribution (orange plot). The vertical line marks the mean with Treatment students slightly higher (more than one half point). When viewed as stacked distributions (both blue), Treatment students showed a slightly more narrow distribution (fewer students in the lower end and a tighter group of students in the higher end) than Control students Following the two distribution graphs, we display a box and whisker box plot. The top of each box (Control and Treatment) represents the score associated with the 75th percentile rank (PR); the middle of the box reflecting the score associated with the 50th PR; the bottom of the box

reflecting the 25th PR. Outlier scores are above and below (the line extending above to the 90th PR and below to the 10th PR) and individuals with data points. The primary interpretation from these plots is that students in the Control group were slightly lower, particularly from the 50th PR and below.

Finally, a bar chart presents the performance of students in New York City who participated (all of them in the Treatment Group). Similar to the larger distributions, these students performed in the upper score range; only a couple scored in the low end with everyone above 10 (of the 15 words) and most between 12-14 words correct.

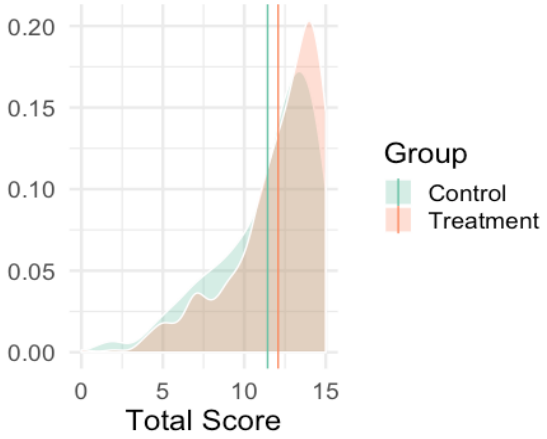

Figure 5. Distribution of scores by treatment and control groups.

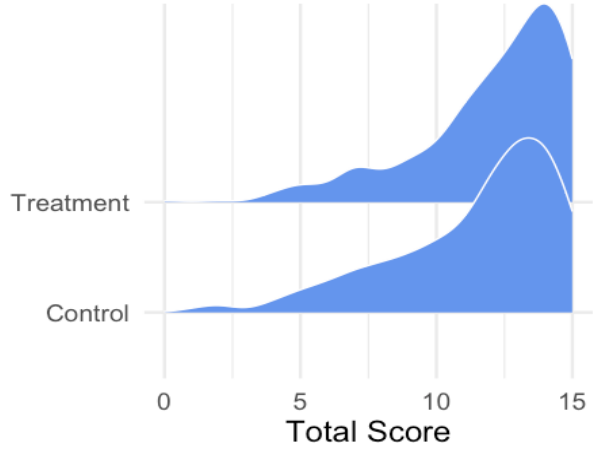

Figure 6. Distribution of scores by treatment and control groups.

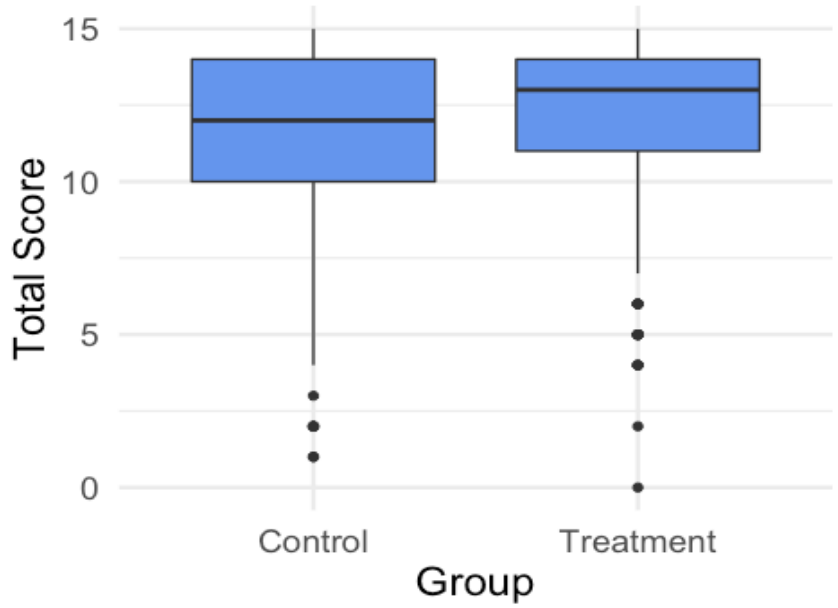

Figure 7. Box and whiskers plot of scores on maze by treatment and control groups.

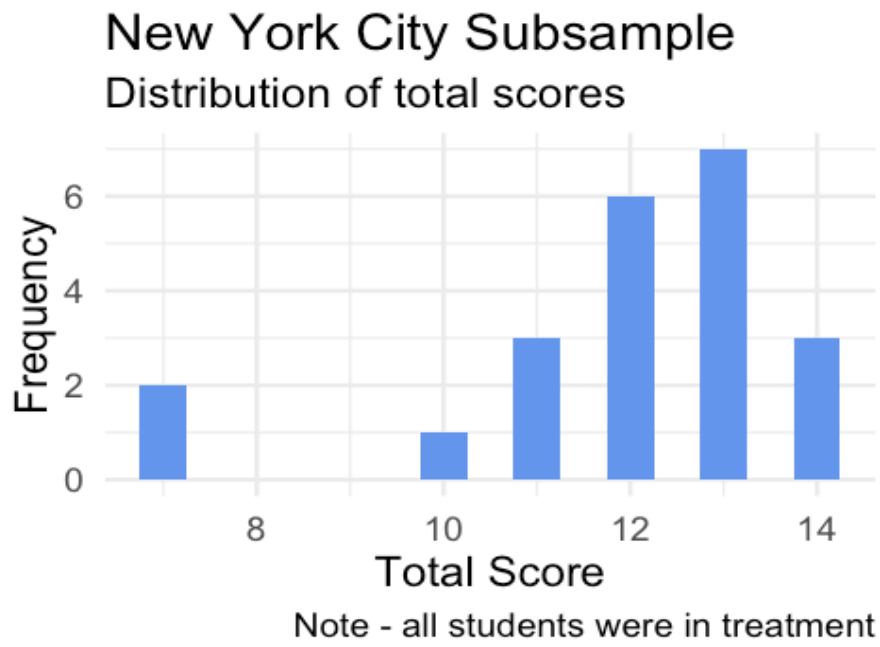

Figure 8. Bar chart of maze scores for the New York city participants.

Given these slight differences between the Treatment Group (A) and the Control Group (B), the question focused on whether it was significant. The regression analysis indicated that, when using the Control Group as a reference: (a) the intercept was about 11 words correct and

(b) the Treatment Group was significantly higher on the maze than the Control Group (with more than an additional half point gained for each point attained by students in the Control group). The probability of being wrong in this conclusion is less than .001, but the variance accounted for represents a relatively small amount (e.g., R2 was low with a value of .012).

Table 3

Regression Table Depicting Significance of Treatment Relative to Control

	(1)
(Intercept)	11.430 ***
	(0.107)
Treatment	0.646 ***
	(0.169)
N	1222
R2	0.012
logLik	-3029.694
AIC	6065.388

**** p < 0.001; ** p < 0.01; * p < 0.05.*

In the next set of analyses (with column 1 repeating the analysis above), time was taken into account as a covariate in columns 2 (time unconstrained) and 3 (time limited to one hour or less). We measured time as the difference between the last opened and the submitted response. Although the model in column 3 indicated a significant effect from the treatment, no effect was present when time was taken into account; the amount of variance explained (R2) did not change.

Table 4

Regression Table Depicting Significance of Treatment (Group A) with Three Models

	(1)	(2)	(3)
(Intercept)	11.430 ***	11.650 ***	11.650 ***
	(0.107)	(0.181)	(0.181)
groupA	0.646 ***	0.641 ***	0.641 ***
	(0.169)	(0.169)	(0.169)
time_taken		-0.032	-0.032
		(0.021)	(0.021)
N	1222	1222	1222
R2	0.012	0.014	0.014
logLik	-3029.694	-3028.552	-3028.552
AIC	6065.388	6065.104	6065.104

**** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.*

In the final set of analyses, each item is analyzed for its functioning: (a) in relation to the distractors and (b) in terms of reliability (consistency). Each plot shows the three options with the correct answer listed first and the two incorrect options listed subsequently with the colored lines being a smoothed fit and the gray line reflecting the actual values. The plot displays the relation between the option being selected (the proportion of students selecting it on the y axis) and their performance on the maze on the x axis (with lower scores on the left and higher scores on the right). Consistently, a relation exists with proportionately more students selecting the correct option as they perform with higher total scores. The displays also show the relative lack

of consistent distraction for the two incorrect options. Finally, the plots reveal a ceiling effect; the maze was relatively easy.

Item Option Analysis

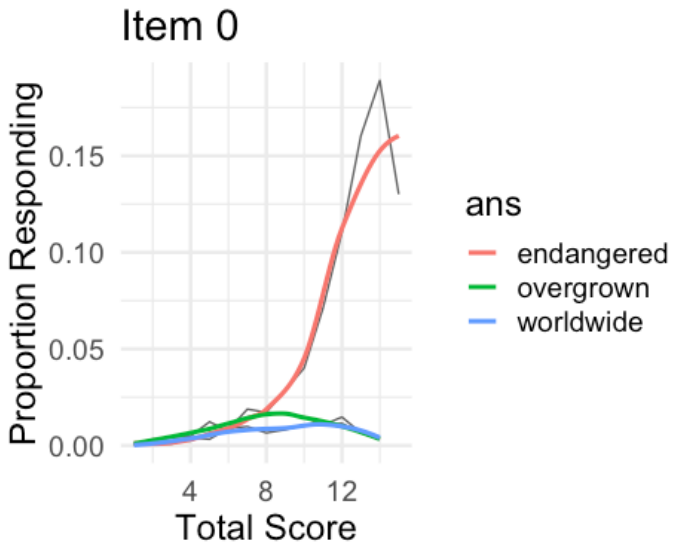

Figure 9. Distribution of proportion correct on total score by option: Item 0.

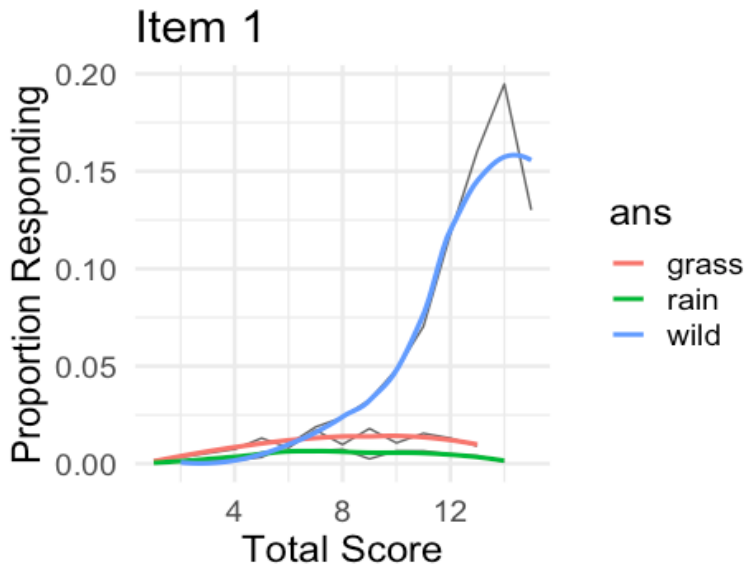

Figure 10. Distribution of proportion correct on total score by option: Item 1.

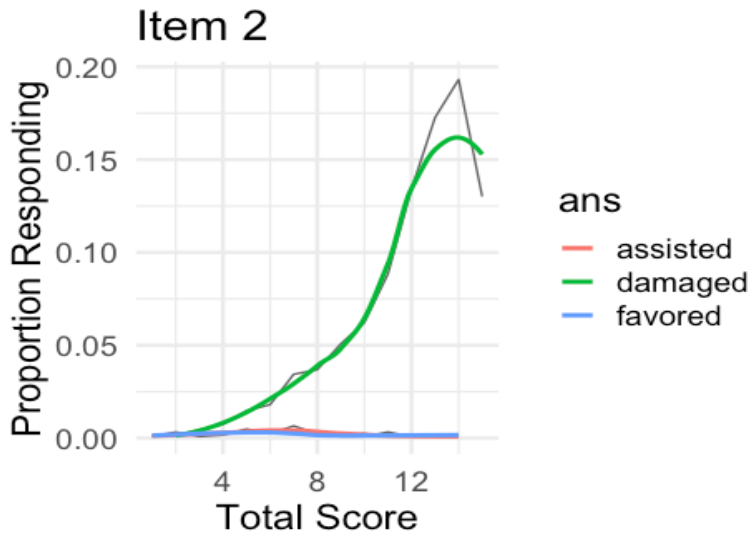

Figure 11. Distribution of proportion correct on total score by option: Item 2.

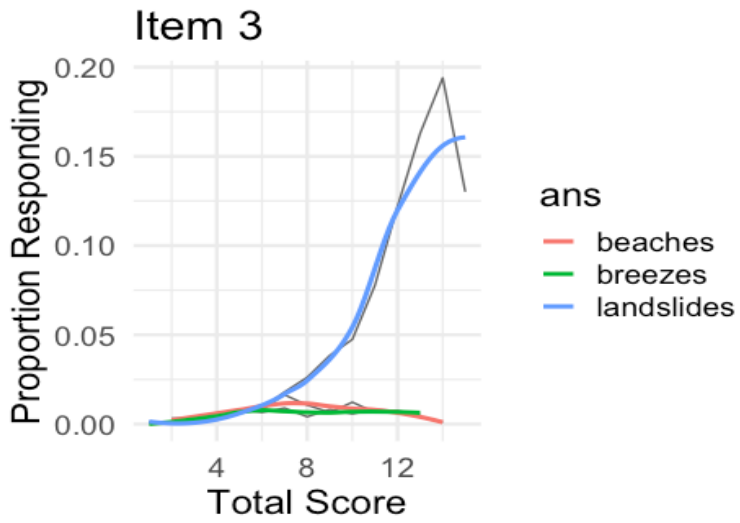

Figure 12. Distribution of proportion correct on total score by option: Item 3.

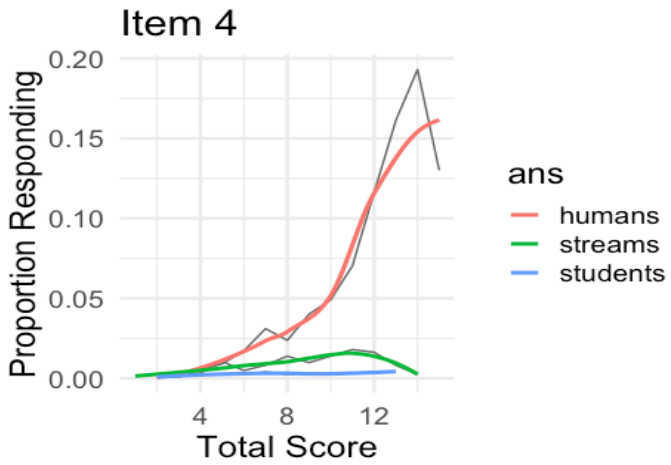

Figure 13. Distribution of proportion correct on total score by option: Item 4.

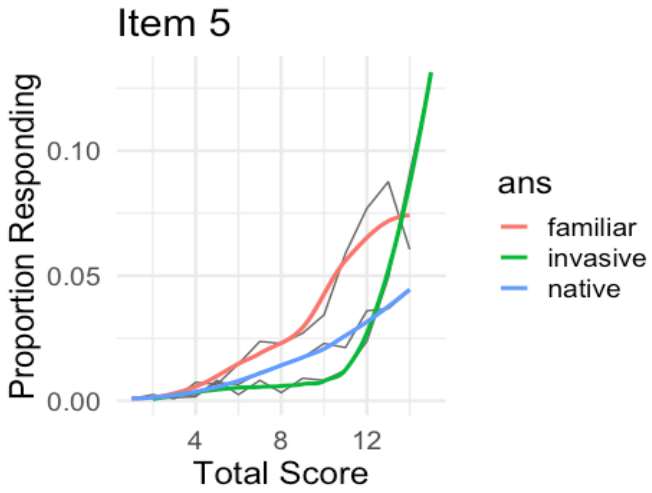

Figure 14. Distribution of proportion correct on total score by option: Item 5.

Figure 15. Distribution of proportion correct on total score by option: Item 6.

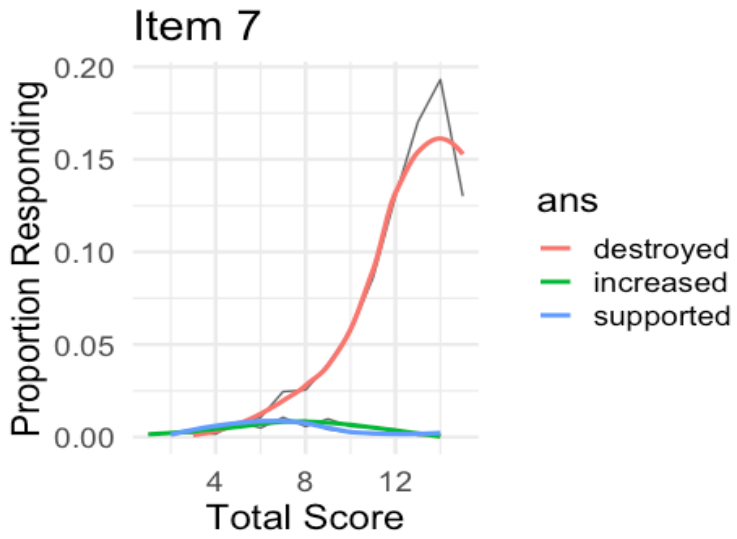

Figure 16. Distribution of proportion correct on total score by option: Item 7.

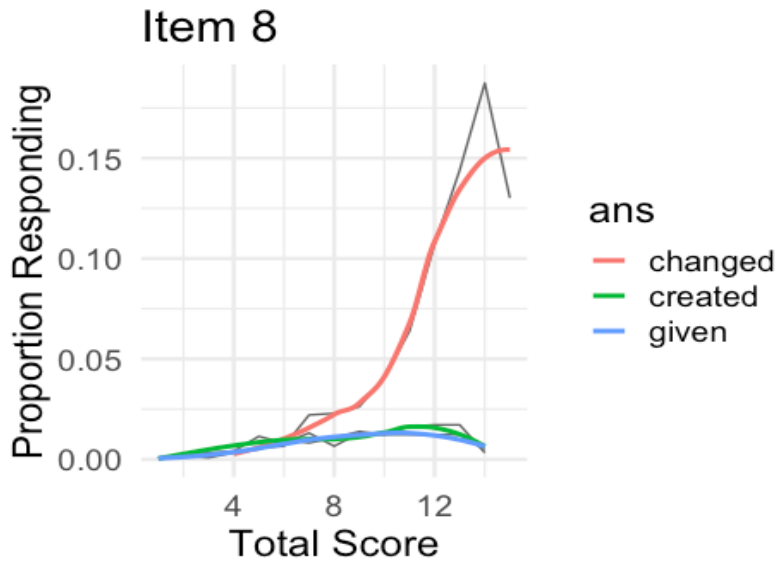

Figure 17. Distribution of proportion correct on total score by option: Item 8.

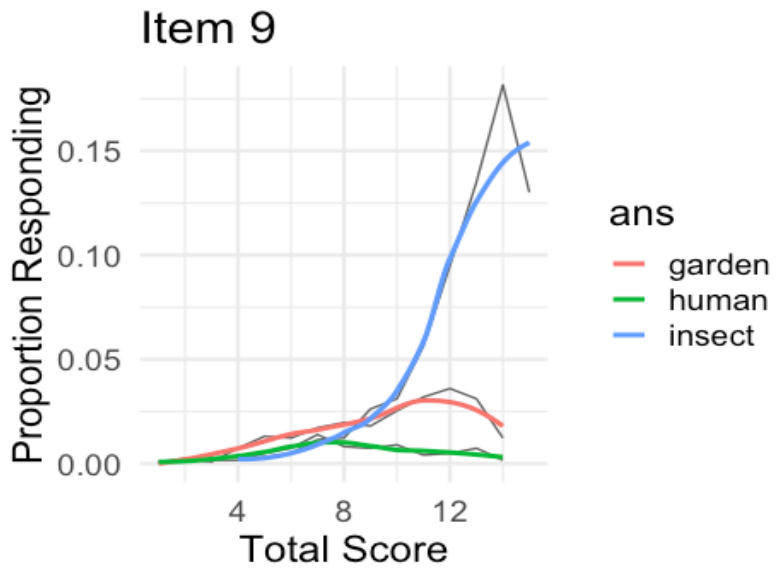

Figure 18. Distribution of proportion correct on total score by option: Item 9.

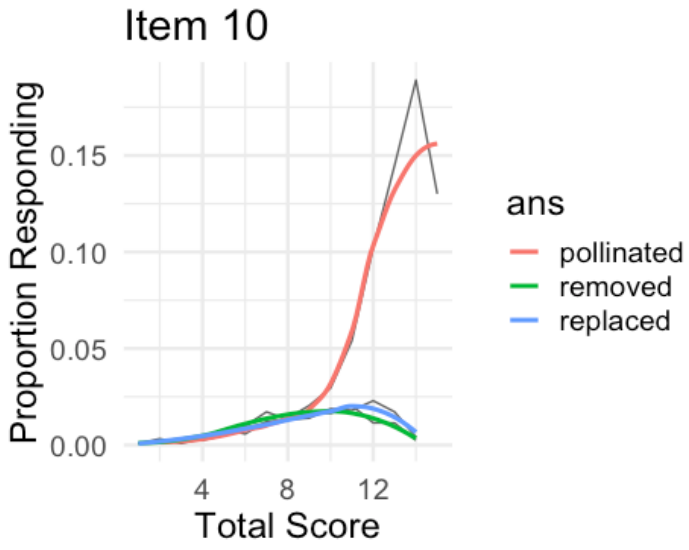

Figure 19. Distribution of proportion correct on total score by option: Item 10.

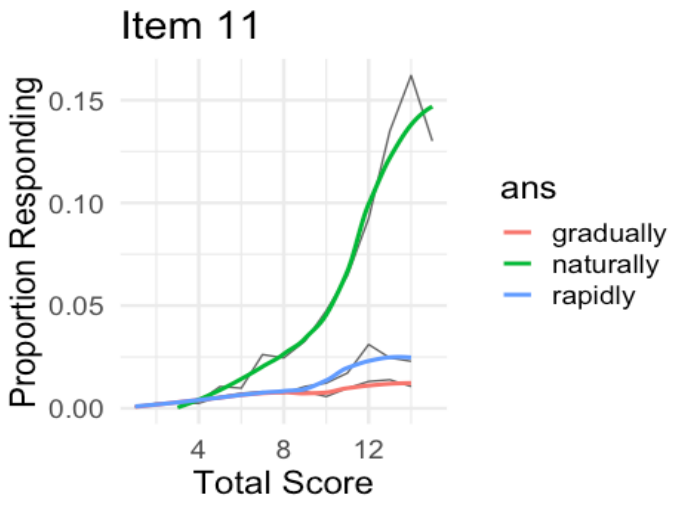

Figure 20. Distribution of proportion correct on total score by option: Item 11.

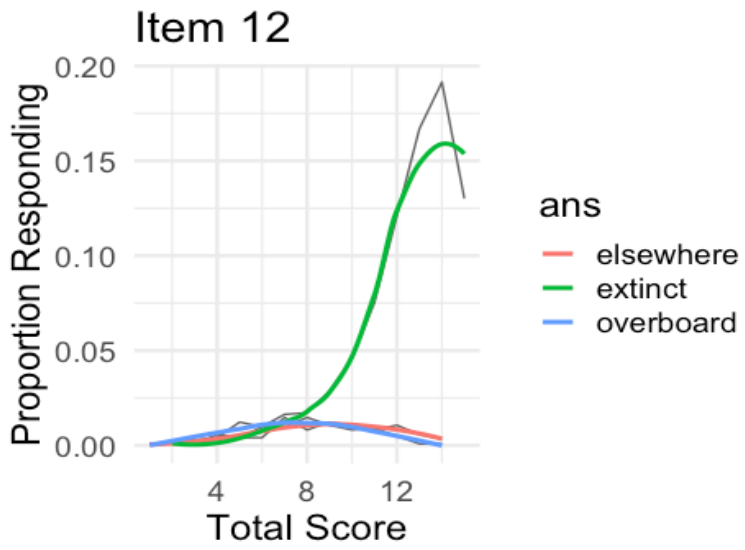

Figure 21. Distribution of proportion correct on total score by option: Item 12.

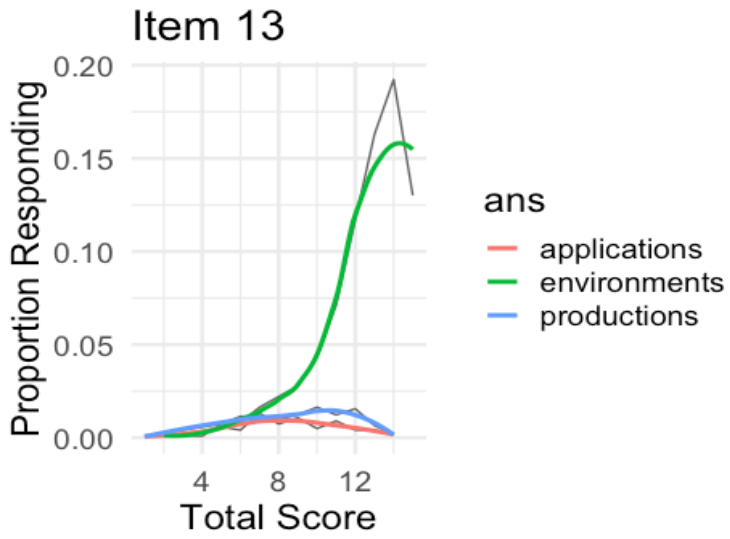

Figure 22. Distribution of proportion correct on total score by option: Item 13.

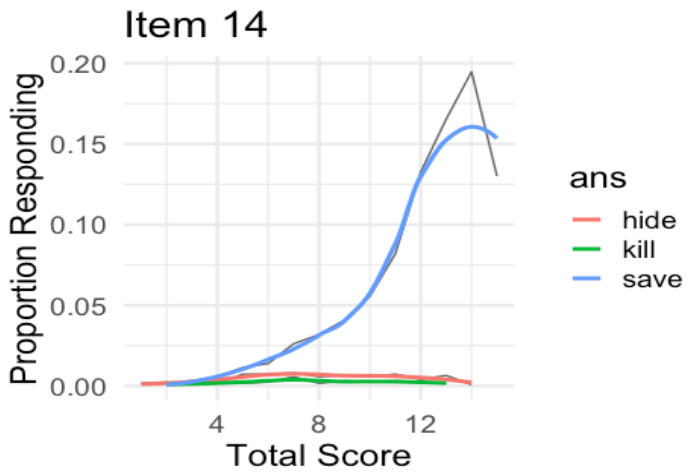

Figure 23. Distribution of proportion correct on total score by option: Item 14.

In the tables below, overall reliability of the maze is first presented, and then reliability (consistency) of each item is listed in raw score values, standardized values, and with the standard error. The maze assessment overall is reasonably reliable overall (.76) and with all items (ranging in the same values).

The final tables present item statistics with the reliability presented as raw score values, standard score values, when being answered correctly, and with the item being dropped; finally,

the average (mean) and standard deviation (sd) are presented. The important interpretation from this table is the high mean for all the items except for item 5.

Reliability Analysis

Table 5

Overall Internal Consistency for Maze

estimate	lower	upper
0.76	0.74	0.78

Table 6

Reliability with Item Removed

	raw_alpha	std.alpha	alpha se
s0	0.74	0.75	0.01
s1	0.74	0.75	0.01
s2	0.75	0.76	0.01
s3	0.74	0.75	0.01
s4	0.75	0.76	0.01
s5	0.76	0.77	0.01
s6	0.74	0.75	0.01
s7	0.74	0.75	0.01
s8	0.75	0.76	0.01
s9	0.74	0.75	0.01
s10	0.74	0.75	0.01
s11	0.77	0.77	0.01
s12	0.74	0.75	0.01
s13	0.74	0.75	0.01
s14	0.75	0.76	0.01

Table 7

Item Statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
s0	1220	0.53	0.53	0.47	0.41	0.79	0.41
s1	1218	0.57	0.57	0.53	0.46	0.82	0.39
s2	1221	0.40	0.45	0.38	0.32	0.95	0.23
s3	1215	0.53	0.54	0.50	0.43	0.84	0.36
s4	1217	0.40	0.41	0.33	0.28	0.85	0.35
s5	1216	0.39	0.35	0.25	0.23	0.35	0.48
s6	1215	0.52	0.52	0.48	0.41	0.83	0.37
s7	1215	0.54	0.56	0.52	0.45	0.88	0.32
s8	1215	0.49	0.48	0.42	0.36	0.76	0.42
s9	1210	0.54	0.51	0.46	0.40	0.70	0.46
s10	1208	0.54	0.51	0.46	0.41	0.72	0.45
s11	1207	0.31	0.29	0.18	0.16	0.75	0.43
s12	1204	0.58	0.58	0.55	0.48	0.82	0.38
s13	1202	0.53	0.53	0.48	0.42	0.81	0.39
s14	1201	0.42	0.45	0.38	0.33	0.91	0.29

Discussion

This study addressed the effects from using AAD to improve student comprehension of brief expository passages. After five days, significant differences appeared between students who consistently read the passages and those who had not read them. This difference was both statistically and clinically significant: With a .01 chance of being incorrect in the claim, students in the treatment group improved over one point on the maze for every point achieved by the control group students. This finding may be a function of several factors.

Certainly, the properties of the statistical analyses need to be considered. For example, the sample size was large which makes such results easier to achieve, providing a high level of power. Furthermore, this finding may need to be qualified by the difference in sample sizes for the treatment and control groups, making the analysis unbalanced.

Nevertheless, the results are encouraging for a pilot study and are consistent with the previous research on reading comprehension by national groups. Both the NRP and NRTAC emphasize the importance of vocabulary in teaching and learning students' comprehension. We emphasized vocabulary in this study because of both the treatment and the measurement of outcomes. AAD consists of well-controlled passages of approximately 300 words that can be analyzed in a conceptual framework with key vocabulary. In the five AADs used as part of the treatment and the maze measure used to document outcomes, the key concepts addressed (a) characteristics of a plant (mostly adjectives with a few nouns), (b) the threats to their survival (mostly verbs with a few adjectives), and (c) sources of the threats (mostly nouns). Thus, when students are reading, they may be organizing the content into both lexical/semantic and grammatical categories. The AADs also reflect components of effective vocabulary instruction summarized by the NRP and NRTAC: more opportunities to be actively engaged in reading, repeated exposure to targeted vocabulary, explicit instruction supplemented by teachers, and follow-up questioning provided by teachers (though neither of these last two strategies were monitored).

Given that the purpose of the study was to both investigate the effects of Article-A-Day and document the use of a maze assessment, this study focused on vocabulary as a key component of both the treatment and the outcome. Although the treatment effectiveness was supported (with students in the AAD [treatment] group outperforming students in the control group), two issues appeared with the maze as an outcome assessment. First, the reliability was moderate, both with the overall measure and with items being individually removed (Tables 5 and 6). Second, the distributions of the individual items plotted with proportion correct for each option as a function of the total score displayed an appropriate curve for the correct option but

flat curves for both distractors. In essence, the distractors did not function appropriately for students with different total scores. This is likely a contributor to the relatively high mean score for each item (other than item 5). It is therefore important that construction of the maze needs to take into account both what key vocabulary words are omitted, but also what words are used in their place (Liu, Sundstrom-Hebert, Ketterlin-Geller, and Tindal, 2008). Given these two qualifications of the maze, support for the AAD appears well documented, particularly given the lack of control over actual instructional strategies in the classrooms and the brief duration of the treatment.

References

Butler, S., et al. (2010). *A Review of the Current Research on Vocabulary Instruction: A Research Synthesis*. National Reading Technical Assistance Center.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21*(1), 45-59.

Fuchs, L. S., & Fuchs, D. (2002). Curriculum-based measurement: Describing competence, enhancing outcomes, evaluating treatment effects, and identifying treatment nonresponders. *Peabody Journal of Education, 77*, 64-84.

Fuchs, L. S., & Fuchs, D. (2011). *Using CBM for progress monitoring in reading*. Retrieved from the National Center on Student Progress Monitoring https://files.eric.ed.gov/fulltext/ED519252.pdf.

Hiebert, E. H. (2005). In pursuit of an effective, efficient vocabulary curriculum for elementary students. In E.H. Hiebert & M.L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 243–263). Mahwah, NJ: Erlbaum.

Hiebert, E. H. (2006, April). *A principled vocabulary curriculum*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. *Practical intervention in the schools series*. New York: Guilford Press.

Liu, K., Sundstrom-Hebert, K., Ketterlin-Geller, L., & Tindal, G. (2008). *Instrument development procedures for silent reading measures* (Technical Report No. 0803). Eugene, OR: Behavioral Research and Teaching, University of Oregon.

McCoy, J., Twyman, T., & Tindal, G. (2006). Usng a concept maze to assess student understanding of secondary-level content. *Assessment for Effective Instruction, 31*(2), 39-50.

Marcotte, A. M., & Hintze, J. M. (2009). Incremental and predictive utility of formative assessment methods of reading comprehension. *Journal of School Psychology, 27*, 315-335.

National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC, U.S. Government Printing Office. NIH Publication No. 00-4769.

Parker, R., Hasbrouck, J. E., & Tindal, G. (1992). The maze as a classroom-based reading measure: Construction methods, reliability, and validity. *The Journal of Special Education, 26*(2), 195-218.

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42*(2), 282-296.

Pierce, R. L., McMaster, K. L., & Deno, S. L. (2010). The effects of using different procedures to score maze measures. *Learning Disabilities Research and Practice, 25*, 151-160.

Rockman et al. (2016). *Measurably improving teacher effectiveness and student achievement in reading comprehension: NYC pilot project for innovative research-based teacher training & guidance*. San Francisco, CA: Rockman et al.

Rockman et al. (2017). *Summary of English language arts outcomes at elementary schools implementing ReadWorks Article-A-Day*. San Francisco, CA: Rockman et al.

Shin, J., Deno, S. L., & Espin, C. (2000). Technical adequacy of the maze task for curriculum based measurement of reading growth. *The Journal of Special Education, 34*, 164-172.

Shin, J., & McMaster, K. (2019). Relations between CBM (oral reading and maze) and reading comprehension on state achievement tests: A meta-analysis. *Journal of School Psychology, 73,* 131-149.

Shinn, M. R., & Shinn, M. M. (2002). *AIMSweb training workbook: Administration and scoring of reading maze for use in general outcome measurement*. Eden Prairie, MN: Edformation.

Note 1: This work would not have been possible without the programming and data preparation work from Jeffrey Fleishhacker, Head of Product at ReadWorks.

Note 2: The following authors from ReadWorks include:

- Susanne Nobles, PhD, Senior Director of Teaching & Learning
- Manjula Raman, Director of Content & Curriculum
- Katy Laird, Director of Teacher & School Engagement

Note 3: The following authors from Behavioral Research and Teaching (BRT) include:

- Daniel Anderson, PhD, Research Assistant Professor
- Gerald Tindal, PhD, Director