

Reading Roots 5th Edition: Reading "Between the Lions": Project Summary and Evaluation

Qiyang Zhang, MEd
Amanda Neitzel, PhD

December 2021



JOHNS HOPKINS
SCHOOL *of* EDUCATION

**Center for Research and
Reform in Education**

Contents

Reading Roots 5th Edition: Reading "Between the Lions": Project Summary and Evaluation	1
Method	1
Data	1
Research Design	2
Participants	2
Measures	3
Analytical Approach	4
Results	4
Overall achievement results	5
Differential achievement results: Baseline achievement	6
Discussion	8
References	11

Reading Roots 5th Edition: Reading "Between the Lions": Project Summary and Evaluation

Success for All Foundation (SFAF) is a non-profit organization that strives to enable every child to read and succeed. Its core value is built upon implementing evidence-based and research-proven programs to enhance students' literacy skills. Sirius Thinking, Ltd., is a world-renowned children's media production company that specializes in combining educational content with animated, memorable, and appealing media experiences to increase students' learning engagement. These two organizations worked together under a Small Business Innovation Research (SBIR) grant to develop a pilot project for Reading "Between the Lions" called Reading Roots 5th Edition (RR5).

Reading "Between the Lions" brings the famous public television series "Between the Lions" into classrooms daily to enhance teachers' instruction. Building upon the proven instructional power of SFAF's original Reading Roots curriculum, RR5 integrates Reading "Between the Lions" as an additional resource. The goal of RR5 is to improve first grade students' reading skills through highly engaging, fast-paced teacher instruction, and partner-based practice activities, built around a synthetic phonics approach to early reading, supplemented by targeted media elements. Simply speaking, RR5 is a multimedia classroom instructional program that leverages technology to engage students and customize literacy instruction with instructional videos and collaborative games.

SFAF and Sirius Thinking contracted with the Center for Research and Reform in Education (CRRE) at Johns Hopkins University in February 2021 to conduct an evaluation of the initial implementation of RR5 and determine the program's impacts on students' reading outcomes.

The study employed a one-group, pre-post correlational study design to address the following research questions:

1. What was the effect of the RR5 curriculum on first grade students' reading achievement and fluency?
2. Did the effects of the RR5 curriculum differ by baseline achievement?

Method

Data

Data used in this study were primarily provided by SFAF. These data included school names, teacher names, pilot start and end dates, pre-pilot and post-pilot Reading Mastery Level, pre-pilot and post-pilot reading fluency measured by the

number of words read correctly in a minute (WCPM). These data were collected during the RR5 intervention from February to May 2021. In addition, data on school location, size, and demographic composition were retrieved from the Common Core of Data (<https://nces.ed.gov/ccd/>). Reading proficiency data were retrieved from each state's report cards (Arizona School Report Cards, 2021; Kansas Report Card, 2021; Louisiana School Report Cards, 2021; Ohio School Report Cards, 2021; Virginia Department of Education, 2021).

Research Design

The study utilized a one-group, pre-post correlational study design, looking at outcomes for students before they participated in the intervention (February, 2021) and after they participated in the intervention (May, 2021). The intervention lasted for 10 to 12 weeks of instruction.

Participants

Participants were 183 first grade students from across the United States. Nine public schools and 17 teachers took part in the intervention. A convenience sampling strategy was used to recruit schools that have collaborated with SFAF before and were willing to conduct the designed intervention. Although these schools were not selected randomly, every first-grade student in all nine selected schools participated in the intervention. Table 1 displays the characteristics of the districts and schools who implemented RR5. These were a diverse set of schools, ranging from rural to urban settings, smaller and larger schools, and varying demographic profiles. These schools also represented a broad spectrum of prior academic achievement, as measured by percent proficient¹ in reading on their state standardized tests.

¹ Reading proficiency was reported as two different types depending on the state's report cards. In some states, such as Ohio, Louisiana and Virginia, district data was available for the percentage of students either on track or proficient in English Language Arts. In other states, such as Kansas and Arizona, proficiency was stated in four levels of ability to understand and use reading skills: level 1 represents limited ability; level 2 represents basic ability; level 3 represents effective ability; and level 4 represents excellent ability. In the former type, the percentage proficient reported was coded directly into the dataset. In the latter type, a sum of level 3 and level 4 indicated reading proficiency. All reading proficiency refers to the school district's performance as a whole. When first-grade academic performance was unavailable, the nearest grade (usually grade 3) was used to replace missing data. The last column, "proficiency level," indicates whether the school district's average reading performance is high (reading proficiency > 50%), average (reading proficiency = 50%), or low (reading proficiency < 50%).

Table 1*School location and demographic information*

School	Locale, State	School Size	Demographic	Reading proficiency
Alliance Early Learning	Town: Fringe, Ohio	284	70.42% White, 10.92% African American, 2.82% Hispanic	67.40%
Prairie View Elementary School	Rural: Fringe, Kansas	426	87.56% Hispanic, 10.56% White, 0.47% African American	38.08%
McKinley STEM Academy	City: Small, Ohio	142	40.85% African American, 38.73% White, 0.70% Hispanic	93.70%
Sharonville Elementary School	Suburb: Large, Ohio	409	31.54% White, 26.41% Hispanic, 24.94% African American	79.30%
East Garfield Elementary School	City: Small, Ohio	319	45.14% African American, 36.36% White, 4.08% Hispanic	93.70%
Pugliese Elementary West	City: Small, Ohio	476	73.74% White, 13.24% African American, 3.36% Hispanic	93.70%
Parks Primary	Town: Fringe, Louisiana	391	58.57% White, 35.81% African American, 1.53% Hispanic	29.00%
Synergy Public School	City: Large, Arizona	442	70.14% Hispanic, 13.57% White, 10.63% African American	23.00%
Windy Gap Elementary School	Rural: Fringe, Virginia	302	88.74% White, 4.64% Hispanic, 1.66% African American	79.00%

Measures

There are two outcomes of interest in this study: reading achievement and fluency. Reading achievement is measured by Reading Mastery Level (Madden & Slavin, 2017). This is part of the standard Reading Roots formal assessment used in all SFAF schools. Classroom teachers conduct this assessment following a consistent protocol. The assessment consists of individually-administered tests of word attack, letter-word identification, fluency, and comprehension, as well as informal teacher monitoring (records of teacher observations of student reading) and formal classroom measures (standards-based tests, Fast Track Phonics assessments). In this way, the Reading

Mastery Level combines both the formal assessment and the classroom indicators into a single outcome.

Reading fluency is routinely collected by SFAF as another dimension to reflect reading skill. Specifically, reading fluency is individually assessed, and measured by words correct per minute (WCPM) as a teacher listens to a student read. WCPM is calculated by taking the total number of words read in one minute and subtracting the number of errors. In this way, both reading speed and accuracy are accounted for.

Analytical Approach

Linear mixed-effects models were used to analyze the association between the RR5 curriculum with reading achievement and fluency, while controlling for school effects. Linear mixed-effects models are able to account for multiple responses from the same students, as well as the clustering of students within groups such as classrooms or schools. Reading achievement and fluency outcomes were analyzed separately. Models controlled for school demographic composition, school size, school achievement, and student baseline reading fluency. All covariates were grand-mean centered to facilitate the interpretation of the model.

Differential effects of RR5 across baseline achievement were also explored. In each teacher's class, the lowest 25% of students in reading fluency were categorized as low baseline achievement. To examine these differential treatment effects, we added interaction terms between time points (e.g., pre- or post-assessment) and each student subgroup. The treatment effect for each group was calculated by summing the overall treatment effect and the differential treatment effect for each subgroup indicated by the interaction term. All analyses² were conducted in Rstudio Version 1.4.1717 (RStudio Team, 2021) using the R project for statistical computing Version 4.1.1 (R Core Team, 2021).

Results

In the data cleaning step, one observation was removed due to missing data in assessments and four observations were removed due to missing school or teacher. The analytic sample therefore included 178 students. Table 2 presents descriptive statistics for participating schools' demographic composition. Across the nine participating schools, schools had an average of 47% White students, 20% Black students, and 22% Hispanic students.

² We cleaned data using *readxl* (Wickham & Bryan, 2019), *janitor* (Firke, 2021), *tidyr* (Wickham, 2021); conducted linear regression analyses using *lmerTest* (Kuznetsova et al., 2017), *car* (Fox & Weisberg, 2019); and produced tables and figures with *flextable* (Gohel, 2021), *officer* (Gohel, 2021), *tableone* (Yoshida & Bartel, 2021), and *ggplot2* (Wickham, 2016).

Table 2*Descriptive statistics for the sample*

Variable	Overall
School Level	
Number of Schools	9
White (mean [SD])	0.47 (0.27)
Black (mean [SD])	0.20 (0.17)
Hispanic (mean [SD])	0.22 (0.33)

Overall achievement results. Table 3 presents linear mixed effects model results for reading achievement (Reading Mastery Level) and reading fluency (WCPM). For both achievement and fluency, the posttest scores were significantly higher than the pretest scores. In fluency, the average student, when controlling for school demographics, school size, and school achievement, scored 26.06 words correct per minute before participating in RR5. Students' score increase is statistically significant after participating in RR5, with a mean posttest score of 46.76 words correct per minute ($ES = +1.19$, $p < .001$). A similar pattern was observed for reading achievement measured by Reading Mastery Level, where the average student scored 4.74 at pretest, and significantly improved their score to 7.44 after RR5 intervention ($ES = +2.09$, $p < .001$).

Table 3*Linear mixed effects model results*

	N	Unadjusted mean	Adjusted mean	Difference in adjusted means	Effect size	p-value
Words Correct Per Minute (WCPM)						
Pretest	178	23.27 (17.42)	26.06			
Posttest	178	43.97 (24.47)	46.76	20.70 (1.29)	+1.19***	<.001
Reading Mastery Level						
Pretest	178	4.92 (1.29)	4.74			
Posttest	178	7.62 (1.60)	7.44	2.70 (0.10)	+2.09***	<.001

Notes. (a) * $p < .05$, ** $p < .01$, *** $p < .001$. (b) The adjusted means came from the mixed model that controlled for school size, school race/ethnicity composition, and school achievement. (c) The numbers in the parentheses are standard deviations for the unadjusted means and standard errors for the difference in the adjusted means. (d) All covariates were grand-mean centered to facilitate the interpretation. (e) Effect sizes are in terms of standard deviations and were calculated as the treatment effect divided by the pretest standard deviation.

Differential achievement results: Baseline achievement. In addition to testing the effect of RR5 on students' reading outcomes, our model also simultaneously tested for differential effects by baseline reading fluency. Table 4 displays differential effects analysis for WCPM. Differential effects by baseline fluency were not observed on the reading fluency outcome; while students with low baseline fluency had smaller gains between pretest and posttest than students with average and high baseline fluency, these differences were not significant. Differential effects were identified for the reading achievement outcome, in that students with low baseline fluency had average gains of 2.20 units while students with average and high baseline fluency had average gains of 2.83 units. This difference in gains was statistically significant ($p < .01$), but both groups made substantial and significant growth when analyzed separately.

Table 4*Differential effects analysis for separate groups*

	N	Unadjusted mean	Adjusted mean	Difference in adjusted means	Effect size	p-value
Reading Mastery Level						
<i>Average & High Baseline Fluency</i>						
Pretest	139	5.03 (1.26)	4.88			
Posttest	139	7.86 (1.47)	7.71	2.83	+2.19***	<.001
<i>Low Baseline Fluency</i>						
Pretest	39	4.54 (1.33)	4.38			
Posttest	39	6.74 (1.77)	6.58	2.20	+1.70***	<.001
Words Correct Per Minute (WCPM)						
<i>Average & High Baseline Fluency</i>						
Pretest	139	26.49 (17.96)	30.37			
Posttest	139	48.04 (24.75)	51.92	21.55	+1.24***	<.001
<i>Low Baseline Fluency</i>						
Pretest	39	11.79 (8.26)	13.90			
Posttest	39	29.49 (16.98)	31.59	17.70	+1.02***	<.001

Notes. (a) *p < .05, **p < .01, ***p < .001. (b) The adjusted means came from the mixed model that controlled for school size, school race/ethnicity composition, and school achievement. (c) The numbers in the parentheses are standard deviations for the unadjusted means and standard errors for the difference in the adjusted means. (d) All covariates were grand-mean centered to facilitate the interpretation. (e) Effect sizes are in terms of standard deviations and were calculated as the treatment effect divided by the full sample pretest standard deviation.

Table 5*Differential effects analysis for interaction terms*

Test	Coefficient	Impact (SE)	df	t-value	p-value
Reading Mastery Level	Post: low_baseline_wcpm	-0.63 (0.24)	176.00	-2.64	0.009**
Words Correct Per Minute (WCPM)	Post: low_baseline_wcpm	-3.57 (3.28)	175.00	-1.09	0.279

Notes. (a) *p < .05, **p < .01, ***p < .001. (b) The model controlled for school size, school race/ethnicity composition, and school achievement. (c) SE = standard error, df = degree of freedom. (d) All covariates were grand-mean centered to facilitate the interpretation.

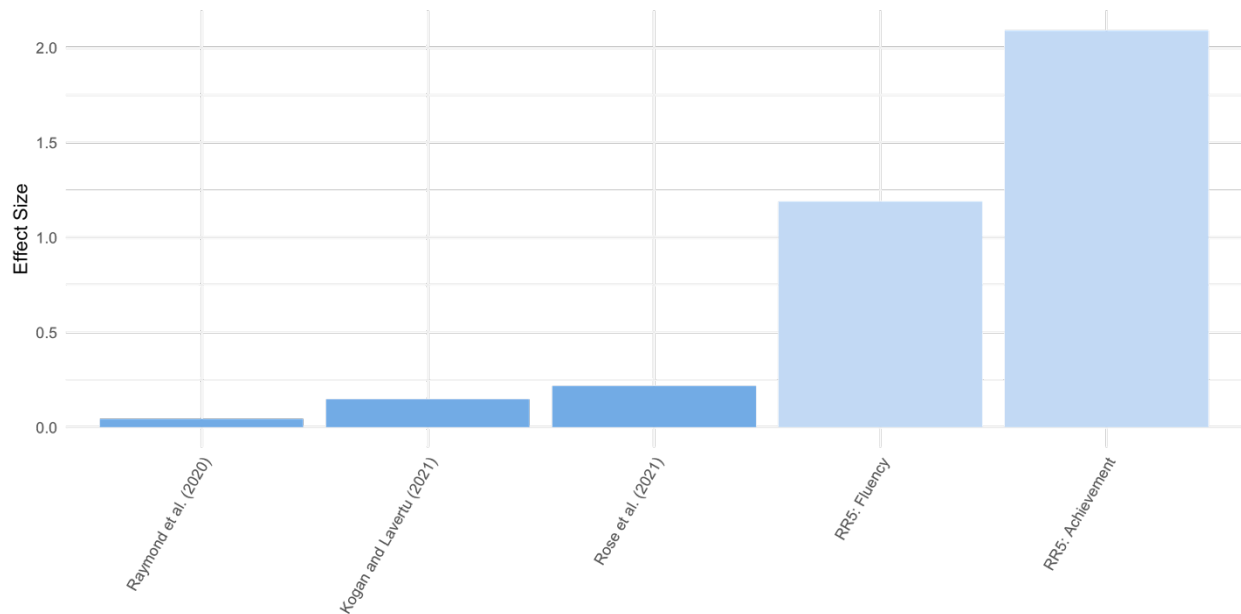
Discussion

The purpose of this study was to examine the efficacy of RR5 on student reading outcomes. Overall, participating students' reading achievement and fluency increased over the period when RR5 was implemented. The intervention took place in the first half of 2021, a time when most U.S. students were still heavily affected by the COVID-19 pandemic physically and emotionally. Without a control group in this pilot study, it can be difficult to interpret these results. However, we looked at students' normal gains during this period. It is important to understand that this sample, while a convenience sample, still represents a "typical" group of first grade students. By comparing the pretest scores of the sample with expected scores from the DIBELS³, we can see that according to DIBELS 8th Edition Benchmark Goals for mid-year first grade students (2020), RR5's average pretests fall into the green range, where students need only core support and are at minimal risk. Students of average and high baseline fluency also fall into the green range while students of low baseline fluency fall into the yellow range, where students need strategic support and are at some risk. This demonstrates that the sample of students participated in RR5 intervention are in the middle ranges since they do not come from extreme ends of achievement, namely the blue and red ranges. Apart from DIBELS, SFAF's Tutoring with the Lightning Squad (2021) has a list of WCPM and respective grade level equivalency levels. Matching RR5's pretest and posttest scores with this reference point, students gained 0.28 increase in grade level, which is approximately 13.44 weeks of learning. This is longer than the actual intervention duration of 10-12 weeks.

Another point of comparison is to consider what the average learning gain is for first grade. According to Hill et al. (2008), from spring of kindergarten to spring of 1st grade, a student will gain, on average, 1.52 SD across an entire school year, compared with the gains of +1.19 and +2.09 SD for less than half a year of RR5.

Since RR5 was implemented during the pandemic, it would generate more insights to compare RR5's results with other studies carried out during this special period. Combing through recent literature, we identified three reference points: Kogan and Lavertu (2021), The Raymond et al. (2020), and Rose et al. (2021). While each of these is from a slightly older sample (2nd and 3rd grade) and on standardized tests, it gives a point of comparison. As depicted in Figure 1, the impacts of RR5 are much larger than typical growth in schools over one semester during this period. This illustrates that the results of RR5 are quite remarkable.

³ In DIBELS' benchmark assessment of words correct, students were required to read three different text passages of their grade level for one minute each (Chapter 9: DIBELS Oral Reading Fluency, 2016). A median score was selected to give a less biased indicator of students' number of words read correctly across passages of different difficulty levels. Therefore, technically, DIBELS Oral Reading Fluency (Words Correct) in the first grade could serve as a comparison group for RR5's WCPM.

Figure 1*Estimates of Learning Gains During COVID Compared with RR5*

This study has several limitations in terms of research design. The first and the main limitation is the absence of a control group. A one-group, pre-post correlational study design cannot strongly support the conclusion that RR5 is an effective intervention. Without demonstrating comparison between treatment and control groups, it is possible to argue that the observed positive effects may be caused by some other confounding variables that were not observed and not included in our regression models—for instance, that students would have grown at that rate without instruction. Future research on RR5 could adopt an experimental randomized controlled trial, the gold standard for effectiveness research.

The second limitation is the short duration of the implementation period: three months. Since RR5 is a year-long curriculum, the effects of a three-month intervention may not be generalizable to the full curriculum. For future work, researchers can look at this intervention across a full school year to investigate whether RR5's positive effects are sustainable in long terms.

The third limitation to consider is that the sampling strategy may limit the results' generalizability. For this pilot study, convenience sampling was used for its low cost and easy accessibility. However, the sample obtained through this method may under- or over-represent the targeted population. The results reported in this report are not generalizable to a wider population of schools and students.

The last limitation is the reliance on a researcher-made measure. Since Reading Mastery Level combines both formal assessments and informal teachers' observations, and was created by the curriculum developers, it is considered as a researcher-developed measure rather than a standardized test. Past studies have demonstrated that researcher-made measures tend to overestimate programs' effect sizes compared to intervention-independent assessments (Cheung & Slavin, 2016; de Boer, Donker; van der Werf, 2014; Wolf, 2021).

Despite these limitations, this pilot study demonstrates the promise of RR5 for raising students' reading outcomes. This discovery has important implications in the post-COVID time. A recent meta-analysis (Storey & Zhang, 2021) synthesized studies measuring COVID learning loss found that younger students, especially those in grades K-2, were affected more than their higher-grade counterparts. Several studies also highlighted the disproportionate levels of learning loss from historically disadvantaged racial and ethnic groups (Amplify, 2021; Dorn et al., 2020; Kogan & Lavertu, 2021). Given the needs to address differentiated learning loss, policymakers are urgently searching for evidence-based proven strategies to address unfinished learning. The RR5 intervention was found to be effective for the first-grade and racially-diverse group of students, therefore, can be considered by practitioners who aim to enhance reading outcomes for young students. In addition, this pilot study also highlights the need for additional studies evaluating the efficacy of RR5 in a larger sample with standardized measures and a more rigorous research design.

References

- Amplify. (2021). Research brief: COVID-19 means more students are not learning to read.
- Arizona School Report Cards (2021, September 24). *Synergy Public School*.
<https://azreportcards.azed.gov/schools/detail/962403>
- Bates D., Mächler M., Bolker B., & Walker S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
<https://doi.org/10.18637/jss.v067.i01>
- Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292.
<https://doi.org/10.3102/0013189X16656615>
- Chapter 9: DIBELS Oral Reading Fluency* (2016) DIBELS Next Assessment Manual. Retrieved November 15, 2021, from
http://www.shastacoe.org/uploaded/Dept/is/2016_3rd_Grade_Testing_Materials.pdf
- DIBELS 8th Edition Benchmark Goals*. (2020, July) University of Oregon. Retrieved November 15, 2021, from
<https://dibels.uoregon.edu/sites/dibels1.uoregon.edu/files/2021-06/DIBELS8thEditionGoals.pdf>
- Donker, A. S., de Boer, H., Kostons, D., Dignath van Ewijk, C. C., & van der Werf, M. P. C. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review*, 11, 1–26.
<https://doi.org/10.1016/j.edurev.2013.11.002>
- Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020). COVID-19 and learning loss—Disparities grow and students need help (Lunch with Lori Series). McKinsey & Company. <https://www.strumpfassociates.com/demo/wp-content/uploads/2021/01/COVID-19-and-learning-loss-disparities-grow-and-students-need-help-V2-1.pdf>
- Firke, S. (2021). *janitor: Simple tools for examining and cleaning dirty data*. R package version 2.1.0. <https://CRAN.R-project.org/package=janitor>
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression*, Third Edition. Thousand Oaks CA: Sage.
<https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Gohel, D. (2021). *flextable: Functions for tabular reporting*. R package version 0.6.8. <https://CRAN.R-project.org/package=flextable>
- Gohel, D. (2021). *officer: Manipulation of Microsoft Word and PowerPoint documents*. R package version 0.4.0. <https://CRAN.R-project.org/package=officer>
- Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical Benchmarks for Interpreting Effect Sizes in Research. *Child Development Perspectives*, 2(3), 172–177. <https://doi.org/10.1111/j.1750-8606.2008.00061.x>
- Kansas Report Card (2021, September 24).
https://ksreportcard.ksde.org/home.aspx?org_no=D0362&rptType=2
- Kogan, V., & Lavertu, S. (2021). *The COVID-19 pandemic and student achievement on Ohio's third-grade English language arts assessment*.

- https://glenn.osu.edu/sites/default/files/2021-09/ODE_ThirdGradeELA_KL_1-27-2021.pdf
- Kuznetsova A., Brockhoff P. B., & Christensen R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26. <https://doi.org/10.18637/jss.v082.i13>
- Louisiana School Report Cards (2021, September 24). *Department of education school & district report cards*. <https://www.louisianabelieves.com/data/reportcards/>
- Madden, N. A., & Slavin, R. E. (2017). Evaluations of technology-assisted small-group tutoring for struggling readers, *Reading & Writing Quarterly*, 33:4, 327-334, DOI: 10.1080/10573569.2016.1255577
- Ohio School Report Cards (2021, September 24). <https://reportcard.education.ohio.gov/home>
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- Raymond, M., Han, C., Goulas, S., Lee, W. F., & Waeiss, C. (2020). *Estimates of Learning Loss in the 2019-2020 School Year*. The Center for Research on Education Outcomes. https://credo.stanford.edu/sites/g/files/sbiybj6481/f/short_brief_on_learning_loss_final_v.3.pdf
- Rose, S., Twist, L., Lord, P., Rutt, S., Badr, K., Hope, C., & Styles, B. (2021). *Impact of school closures and subsequent support strategies on attainment and socio-emotional wellbeing in key stage 1: Interim paper 1*. https://educationendowmentfoundation.org.uk/public/files/Publications/Covid-19_Resources/Impact_of_school_closures_KS1_interim_findings_paper_-_Jan_2021.pdf
- RStudio Team (2021). *RStudio: Integrated development environment for r*. RStudio, PBC, Boston, MA. <http://www.rstudio.com/>
- Storey, N., & Zhang, Q. (2021, September 10). A Meta-analysis of COVID Learning Loss. <https://doi.org/10.35542/osf.io/qekw2>
- The Center for Research on Education Outcomes. (2020). *Estimates of learning loss in the 2019-2020 school year*. https://credo.stanford.edu/sites/g/files/sbiybj6481/f/short_brief_on_learning_loss_final_v.3.pdf
- Tutoring with the Lightning Squad (2021). Success for All Foundation.
- Virginia Department of Education (2021, September 24). *Franklin County Public Schools*. <https://schoolquality.virginia.gov/divisions/franklin-county-public-schools#desktopTabs-2>
- Wickham, H. (2016) *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
- Wickham, H. (2021). *tidyr: Tidy messy data*. R package version 1.1.3. <https://CRAN.R-project.org/package=tidyr>
- Wickham, H., & Bryan, J. (2019). *readxl: Read Excel files*. R package version 1.3.1. <https://CRAN.R-project.org/package=readxl>

- Wolf, R. (2021) *Average differences in effect sizes by outcome measure type* [Conference presentation.] SREE 2021 Conference, Virtual.
https://sree.confex.com/sree/2021/recordingredirect.cgi/oid/Recording10/session1267_0.mp4
- Yoshida, K., & Bartel, A. (2021). *tableone: Create 'table 1' to describe baseline characteristics with or without propensity score weights. R package version 0.13.0.* <https://CRAN.R-project.org/package=tableone>