

Summary of Reading by Third Evaluation (Year 1)

Prepared by

Jonathan Rauh, PhD
VP, Evaluation & Public Policy
Trident United Way
6296 Rivers Ave.
North Charleston, SC 29406
(843) 740-7730
jrauh@tuw.org



Trident United Way

Background

In the Spring of 2017 Trident United Way submitted a request for proposal (RFP) to the four school districts in the Tri-County area: Berkeley County, Charleston, Dorchester Two and Dorchester Four. The intent of this RFP was to fund a three-year demonstration project aimed at improving third grade reading proficiency. In April 2017 TUW received two proposals from the four school districts all intending to contract with the Lastinger Center at the University of Florida, Zucker School of Education. The University of Florida Lastinger Initiative (UFLI) is an ongoing effort by UF faculty and students to improve literacy outcomes for struggling readers. UFLI began in 1998 as a tutoring model for struggling beginning readers. After a decade of results based primarily on teacher interviews and difference of means comparisons, the Lastinger Initiative now encompasses a range of literacy projects from K-12, centered primarily on teacher professional development.

Three school districts, Berkeley, Dorchester Two and Dorchester Four submitted a joint proposal to TUW that utilized the standard UFLI intervention model. This model begins by identifying struggling readers in grades K-2. Utilizing the intervention in K-2, the goal is to have students reading on the third-grade level by the end of grade two. This intervention instructs teachers in a specific pedagogical approach to be used first one-on-one with struggling readers, then in a small-group setting (3-4 students). Charleston County School District submitted a proposal aimed at working with students in three and four-year-old pre-Kindergarten courses. This proposal did not seek the traditional UFLI intervention; rather it sought to engage in an early literacy content clinic and a community of practice workshop. This form of intervention is a scaled version of

the community of practice intervention typically provided by Lastinger to principals, combined with a more in-depth focus on the early literacy training provided in the traditional UFLI model.

Teacher Satisfaction

All of the teachers surveyed stated that they were either satisfied or very satisfied with the training they received and with the implementation of the program. A common refrain was that their ability to be in control of their students' learning and their students' growth was motivational for them. One example of this that was provided recounted an English Language Learner student that began at level A-1 (non-reader) and in three months was at a DRA level 10 (mid-first grade capability).

Teachers were initially reluctant towards the intervention but over the course of the first nine-weeks of school bought into the technique and began to promote the Lastinger Initiative to their colleagues – so much so that requested attendance for the year two training outstripped the available seats. Teachers formed a community of practice around the Lastinger Intervention. In discussions with District staff and with the Lastinger Trainers, both recounted the support that teachers provided to each other and how the level of buy-in and pride in performance increased among the Lastinger trained teachers.

Fidelity of Implementation

The fidelity with which the teachers implemented the trainings in their classes was assessed using both self-evaluation and peer evaluations. Self-evaluations occurred on a semi-monthly basis using a 19 question (binomial – yes/no) instrument created by Lastinger to address the teacher's use of the UFLI model. The same instrument was

used by peers to evaluate the teachers as well. Peer evaluations were conducted quarterly with one evaluation by another Lastinger-trained teacher and one by an Interventionist. It soon became apparent that this method of addressing fidelity of implementation did not provide useful feedback because teachers did not behave the way they normally would when not being monitored. Additionally, teachers stated that they felt as though they were being inspected for errors rather than provided instruction for improvement. To address this, the Lastinger trainers determined that it would be most useful to engage in less intrusive monitoring sessions. This means the teachers were recoded. This was typically done using a smartphone camera and the lesson was uploaded to a shared and encrypted website. The Lastinger trainers then provided feedback to the teachers and offered suggestions for improvement, modifications of techniques, and general feedback on the lessons.

General Findings

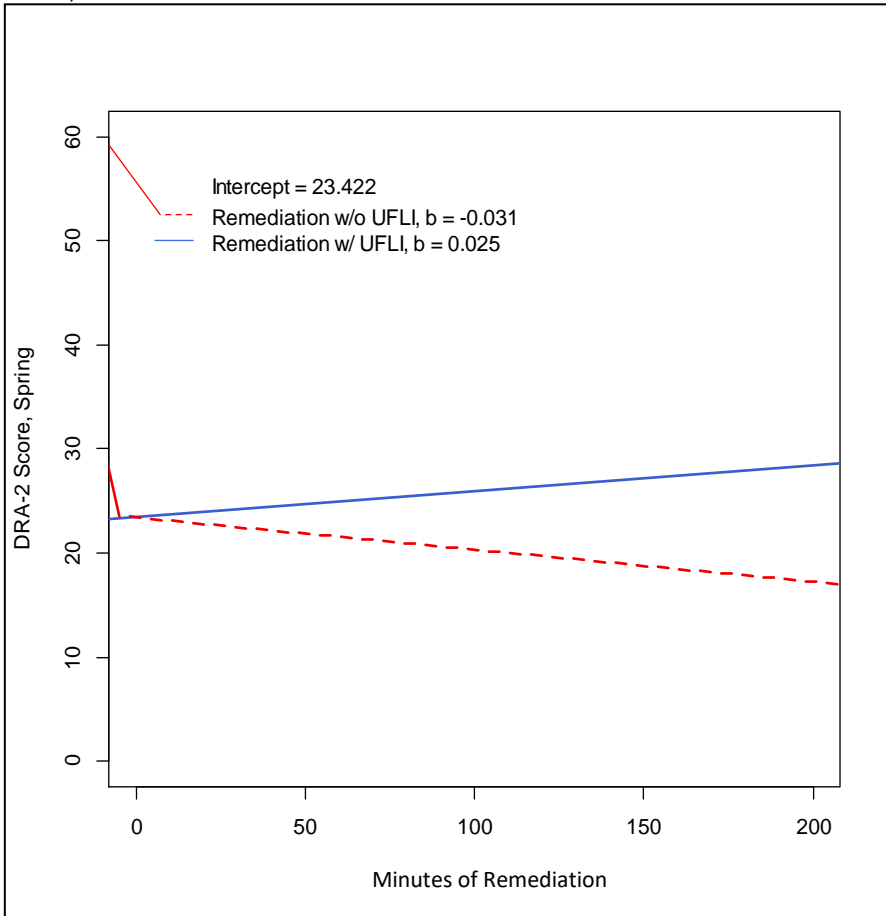
Each of the districts met the required training and reporting requirements. Student growth coefficients equate to between 0.025¹ and 0.445 additional grade-levels of growth on the Developmental Reading Assessment – 2nd Edition (DRA-2). These coefficients are effects on other remediation efforts and a student's prior performance respectively. This corresponds to approximately an additional half-year of growth, e.g. if a 2nd grade student whose prior performance predicted that they should be reading at mid-year 1st grade received the intervention, then they would end up reading at Spring 1st grade/Fall 2nd grade level. There were no statistically significant effects for

¹ This coefficient is based on performance of remediation students in second grade only, e.g. Dorchester District Two's intervention model.

Dorchester Four given that performance among non-remediation students is still relatively low.

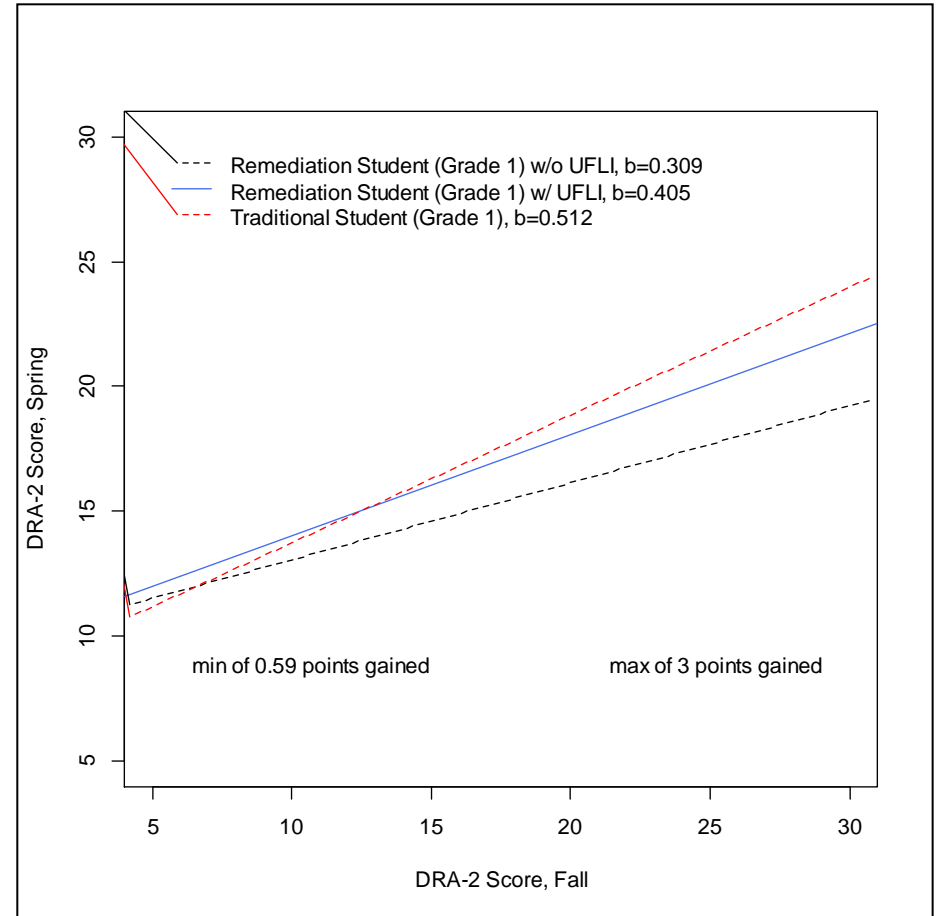
In CCSD the early literacy and community of practice model showed corresponding increases in the probability that students were in their upper performance quartiles in Sound Recognition and Object Naming of 2.5% and 17%. This is in conjunction with decreases in the probability that students would be in the bottom quartile of performance of 16% and 23% respectively.

Figure E1. DRA-2, Second Graders & UFLI Remedial Students (DD2) (BCSD)



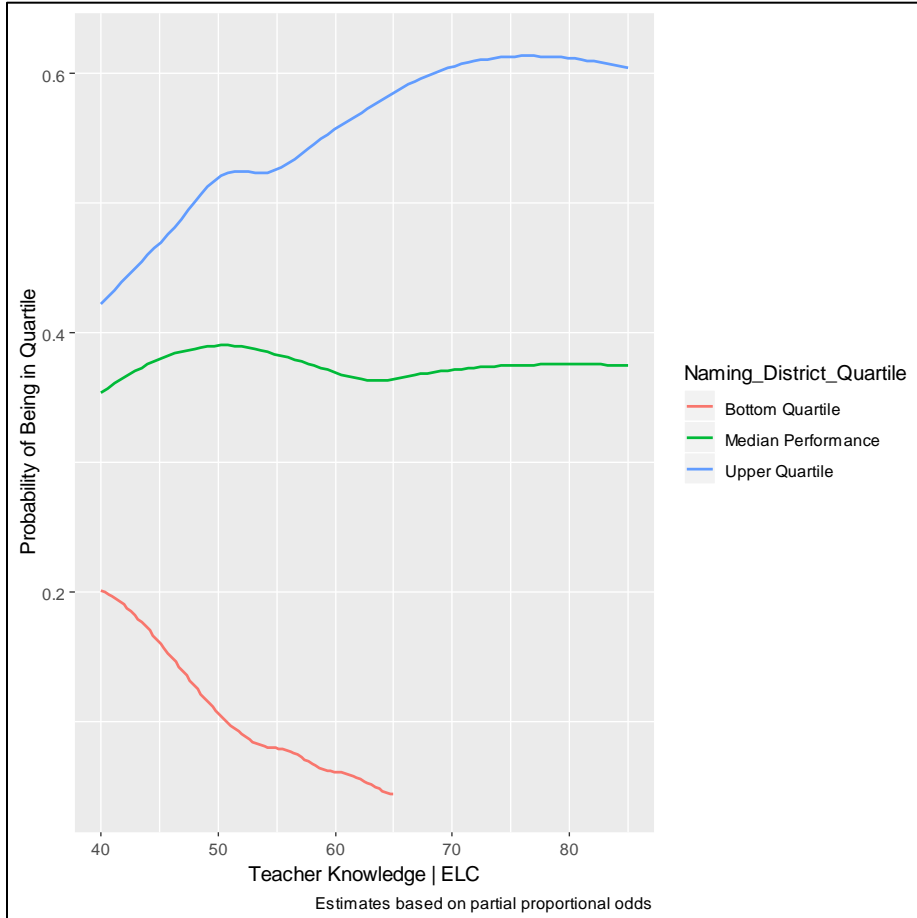
- UFLI has a multiplicative effect on other remediation activities
- Without UFLI students lose 0.031 Spring points for every minute/week of remediation
- With UFLI, students gain 0.025 Spring point for every minute/week of remediation.
- Average minutes/week = 148, e.g. an 8.29 point difference.

Figure E2. DRA-2 Performance for UFLI Student & Average Student



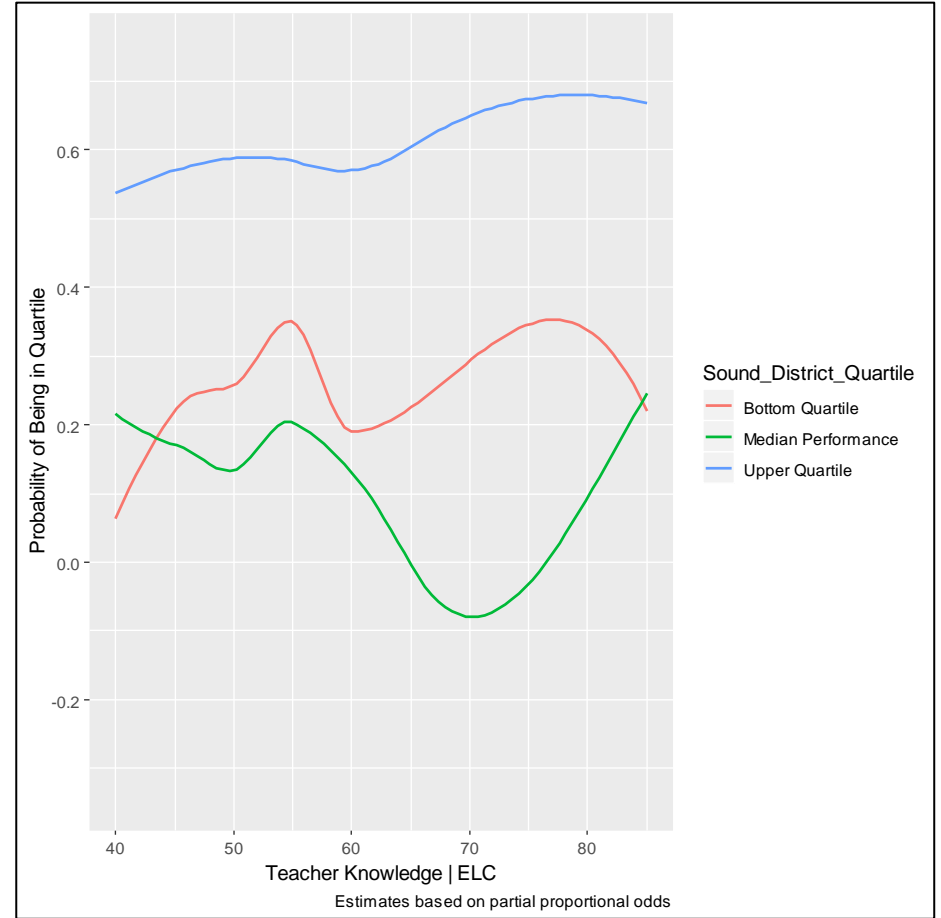
- Fall scores are strongest predictor of Spring scores
- Without UFLI, remediation students gain 0.309 Spring points for every Fall point.
- With UFLI, remediation students gain 0.405 Spring point for every minute/week of remediation.
- This means up to an additional 3 points or ½ grade level due to UFLI

Figure E3. Probability of Being at Level on Naming Scores



- As teacher knowledge increases the probability that a student is at Upper Quartile increases by 17%
- There is also a decrease in the probability that students will be at Lower Quartile or Median of 16%, e.g. even student at the bottom quartile are progressing into the top quartile.

Figure E3. Probability of being at level on Naming Scores



- As teacher knowledge increases the probability that a student is at Upper Quartile increases by 2.5%
- There is a 23% decrease in students being in the Lowest Quartile and a 9.4% increase in students being in the Median.
 - *In general, the largest movement is one level up.*

READING BY THIRD: YEAR TWO EXECUTIVE SUMMARY

General Results

Berkeley County School District

Berkeley County served 69 students across three schools utilizing 25 teachers. Berkeley chose to focus on the lowest performing traditional students as opposed to remedial students. Effects of RB3 are then estimated based on whether the student was in the RB3 intervention as compared to their traditional and remedial peers. Control variables include prior performance, poverty status, parental education level, race/ethnicity, and teacher and school effects. Coarsened Exact Matching procedures were utilized to approximate effects from a randomized experimental design.

On average, RB3 students scored 1.592 points higher than non-RB3 students. This accounts for an additional 64% of growth in Kindergarten, 38.4% of growth in Grade 1 and 16% of growth in Grade 2. As grade level increases, there is a corresponding increase in the baseline score of approximately half a point. Students in remediation tend to score lower than their peers who are not in remediation. Specifically, for every one minutes increase in remediation, there is a corresponding decrease in scores of 0.022 points. With an average of 160 minutes this accounts for approximately 3 points of loss on average. RB3 students' baseline score is, on average, 1.592 points higher than non-RB3 students, however there is not effect from RB3 based on Fall Scores.

Figure E1. Effects of RB3 in BCSD

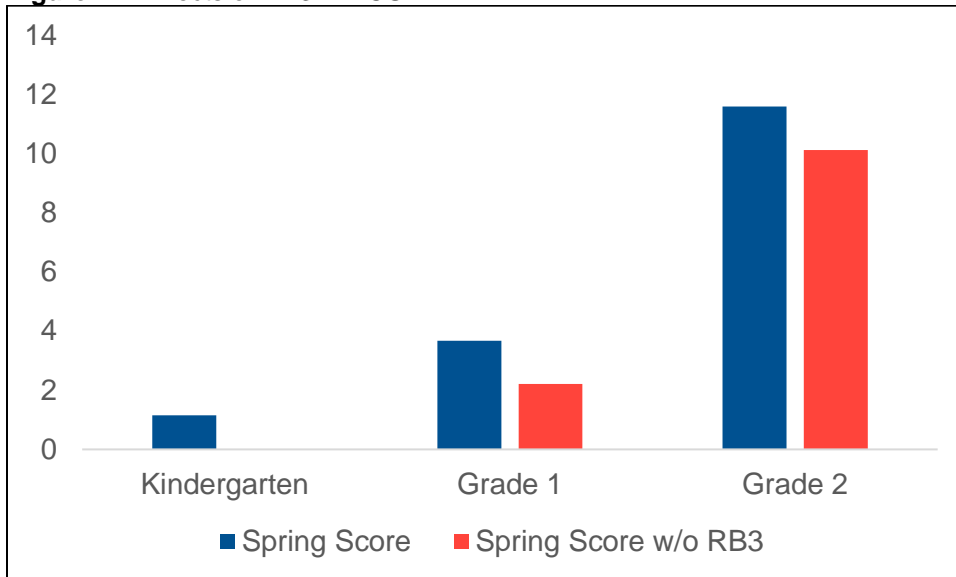
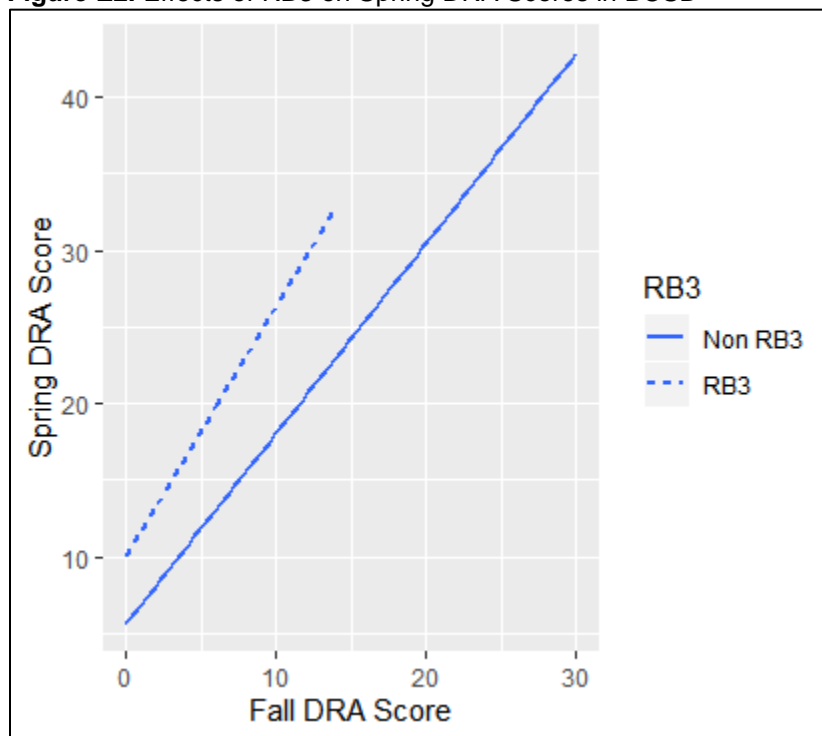


Figure E2. Effects of RB3 on Spring DRA Scores in BCSD



Charleston County School District

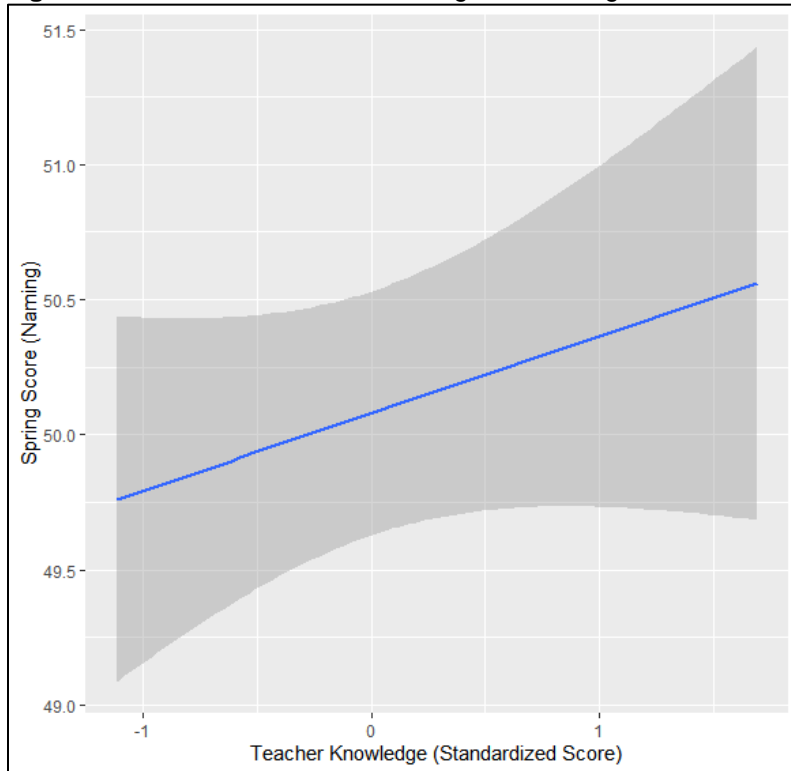
Charleston County served 172 students in 2018-19 across six schools utilizing 11 teachers. CCSD focused on the pre-K component, i.e. pre-literacy, utilizing a theory that

affecting early literacy skills will positively impact student performance in third grade.

Given this approach, all students received instruction from a UFLI trained teacher and were not broken into specific training cohorts in the same manner as the traditional UFLI/RB3 implementation in Berkeley, Dorchester 2 and Dorchester 4.

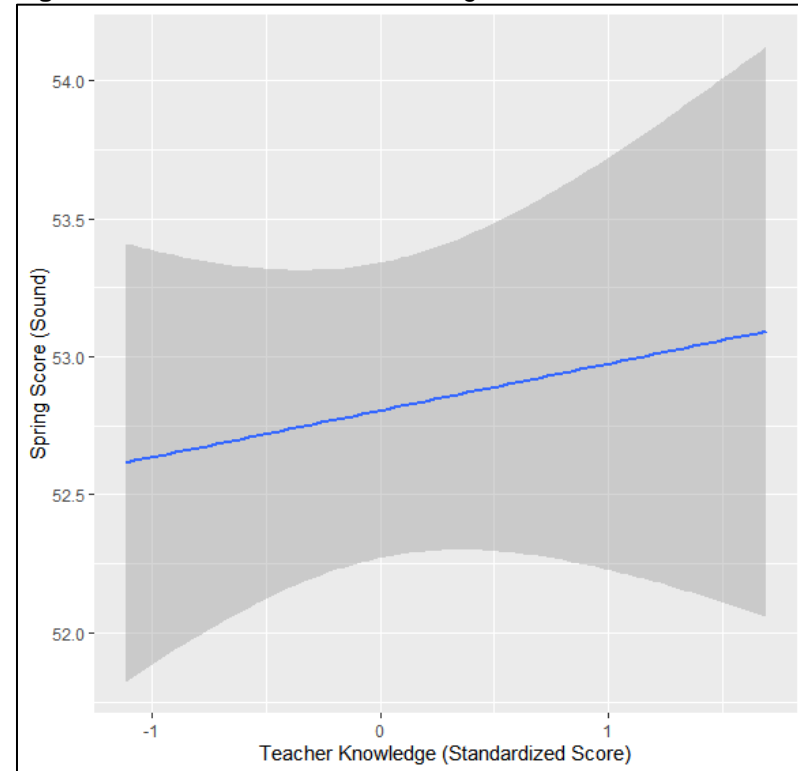
For the Naming portion this means that for every one-standard deviation increase in teacher knowledge following the UFLI training, there is a corresponding increase in Spring scores of 0.442 points. Similarly, for every one-standard deviation increase in knowledge gain there is a corresponding increase in Sound Scores of 0.567 points. This produces a max effect for Naming of 1.226 points and a max effect for Sound of 1.573 points. With an average difference of scores from Fall to Spring in Naming of 4 points and an average difference in Sound scores of 7 points this means that the maximum variation explained by changes in teacher ability given UFLI is 33.75% for Naming and 22.8% for Sound recognition. Although the coefficients for Sound scores were larger than those of Fall Scores, the prior score still predicts the largest share of variation with a total effect change of 1.998 points for Naming (55%) and 1.812 (30%) for Sound. In terms of other effects, as a student's behavioral risk indicator increased there was a corresponding decrease in scores of approximately 0.07 points.

Figure E3. Effect of Teacher Knowledge on Naming Scores



One sd in teacher knowledge gained = 0.567 points student growth

Figure E4. Effect of Teacher Knowledge on Sound Scores



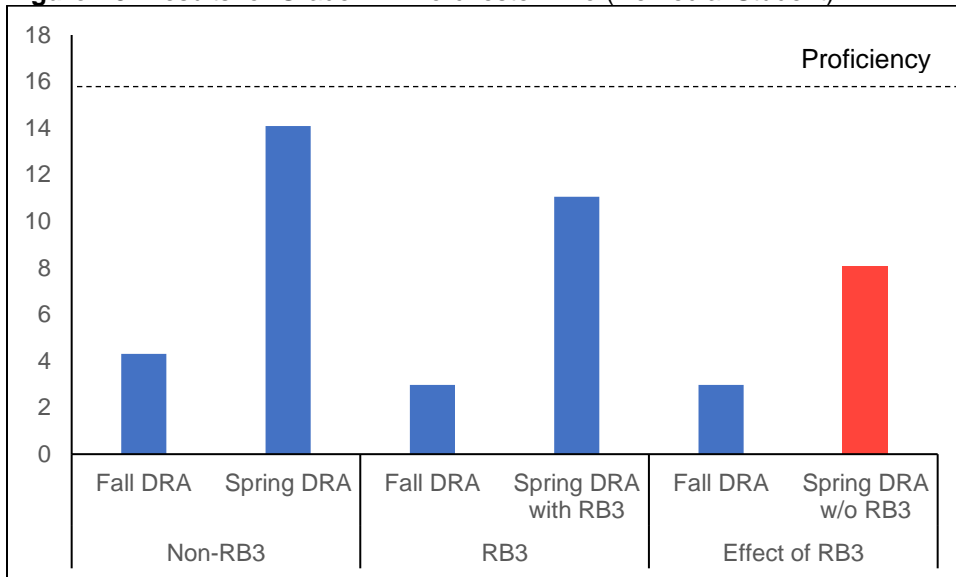
One sd in teacher knowledge gained = 0.442 points student growth

Dorchester District Two

Dorchester District Two served 167 students in RB3 interventions across three schools utilizing 61 teachers in three grades (K-2). Dorchester Two focused on students who were in remediation but scoring on the upper end for remedial students and traditional students scoring on the lower end relative to their traditional peers. The effect of RB3 is assessed by examining students who were in RB3 versus those who were not in RB3 based on remediation status and other relevant factors including prior performance, IEP status, race/ethnicity, poverty status, grade level and teacher and school effects.

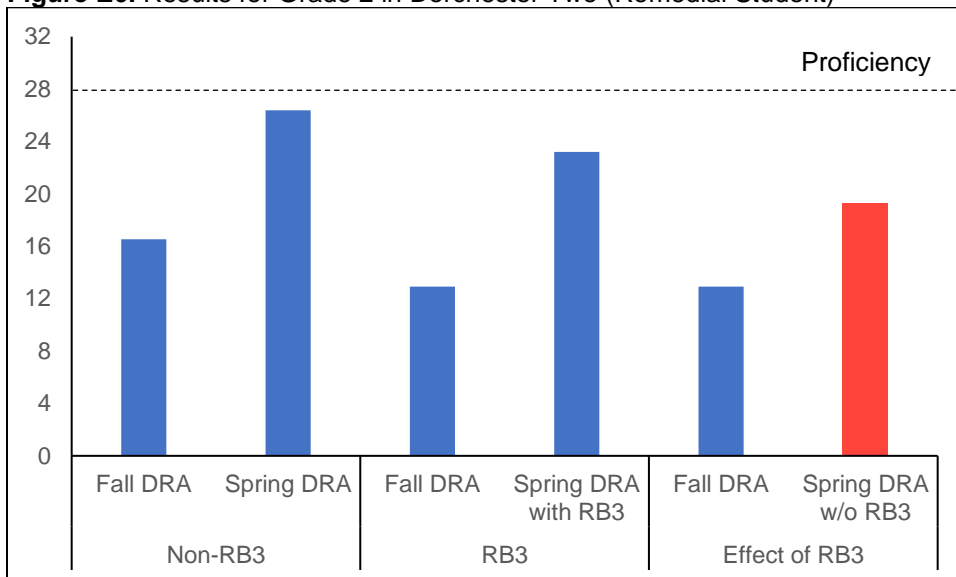
the baseline performance for students in RB3 (remediation students) was not significantly different than traditional student. Based on the fall score coefficients, traditional students show a 0.923 point increase in their spring score for every one point increase in their fall scores while students in RB3 show a 0.213 point increase. If we consider a model without effects from RB3 and use the likelihood ratio test to compare it to the model with RB3, we see that the model with RB3 explains a significantly larger share of variation, e.g. a likelihood ratio of 12.462 increases the probability of a score increase by more than 50%. Turning to the effect sizes within each grade based on movement from the average score to the max score and the corresponding DRA cutoffs (table N) , we see 14% of one year's growth in Kindergarten, 46% of one year's growth in Grade 1, and 30% of one year's growth in Grade 2.

Figure E5. Results for Grade 1 in Dorchester Two (Remedial Student)



RB3 accounted for 46% of one year's growth in DRA for First Grade Students

Figure E6. Results for Grade 2 in Dorchester Two (Remedial Student)



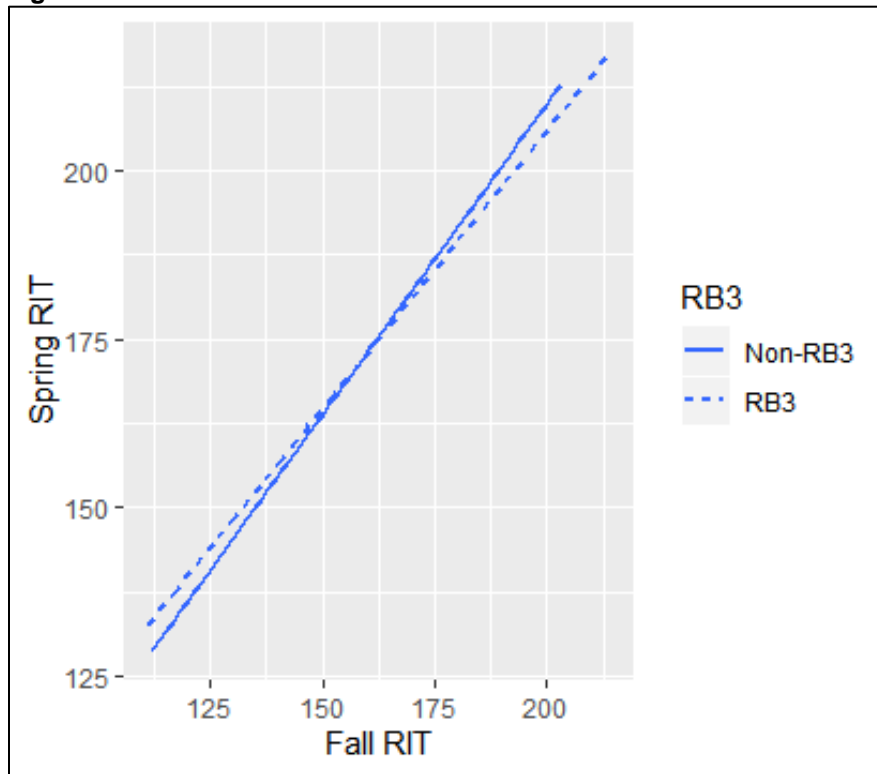
RB3 accounted for 30% one year's growth in DRA for Second Grade Students

Dorchester District Four

Dorchester Four served 238 students across three schools utilizing four teachers in three grades. Dorchester Four did not show a discernable pattern of student selection into RB3 based upon the data (RIT scores). Specifically, both low and high performing students were selected into RB3 while some low performing students (even marginally lower or higher than selected low performers) were not selected into RB3; this raises concerns over the generalizability of the data but also provides potential administrative avenues to examine how students into Dorchester Four are selected into other remediation program.

Fall scores continue to be the strongest predictors of Spring performance, with every one-point of Fall performance contributing 0.671 points to Spring performance. Students in RB3 had baseline scores that were approximately 15 points higher than their non-RB3 peers. However, while non-RB3 students made 0.671 points of growth, RB3 students made approximately 0.58 points which serves to level off results over Fall RIT scores. This translates to approximately 7% additional growth that would not have occurred were RB3 not present. One should interpret these results cautiously given the lack of a data-driven selection criteria in Dorchester Four.

Figure E7. Effects of RB3 on RIT Scores in Dorchester Four



Limited effects in DD4 due to student selection challenges.

Reading by Third: Year Three Executive Summary

General Findings

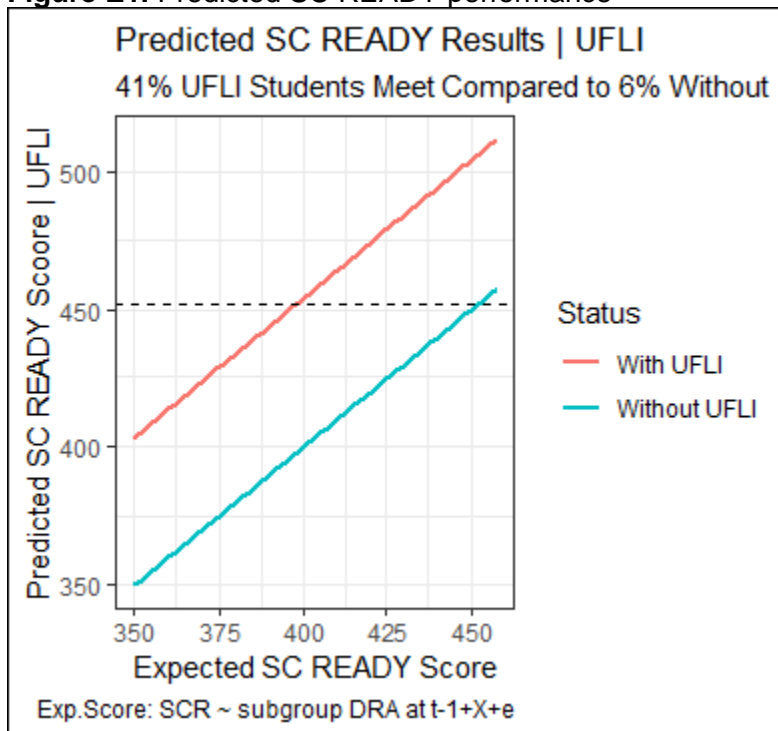
Over a three-year period a total of 4,775 students were taught by a UFLI trained teacher. This is comprised of 1,054 (Berkeley) 1,515 (Dorchester Two), 1,619, and 567 (Charleston). Of the 4,775 students taught, 729 students were struggling readers targeted for UFLI. The selection of students into UFLI was different between districts and the districts' strategies for implementation varied. Dorchester Two focused exclusively on struggling readers, in particular those between the 25th and 50th percentile in terms of fall performance. This selection process was reliable 89% of the time, e.g. if a student were randomly selected from all UFLI students in a DD2 school then 89% of the time that student would be indistinguishable from any other UFLI student in terms of prior performance, race, or poverty. Berkeley focused on students between the 25th and 60th percentile, e.g. both struggling readers and those who are on the cusp of being struggling. The reliability in Berkeley was 92%. Dorchester Four did not have a reliable selection mechanism, neither for identifying students into UFLI nor for identifying students in remediation, e.g. students who had initially scored in the 75th percentile in ELA were identified as needing remediation for ELA. This was addressed with the DD4 administration and selection was adjusted such that we saw 84% reliability in 2019-20.

In terms of performance over a three-year period, the average predicted SC READY performance for UFLI Students was 440 (Approaches). This would place these students, on average, approximately 12 points away from meeting grade level performance expectations whereas before they would have been 61 points away from reaching grade level performance expectations. This is a 49-point improvement, or one-

half of one standard deviation improvement. Without UFLI, these same students are predicted to have scored, on average, 391 (Also Approaches). It should be noted that the average expected performance for traditional students in these same schools is 437, e.g. UFLI scores are on average higher than those of traditional students.

In terms of how this average performance is distributed, we see significantly increased probabilities of scoring Met for students who started with UFLI in grade 1. While the 441 average performance is still within the range of “Approaches,” the distribution of student performance shows that 41% of students are likely to have scored Met whereas if we remove the effects of UFLI we would only expect 6% of students to have scored Met (see figure E1).

Figure E1. Predicted SC READY performance



The effects of UFLI were most pronounced in Berkeley (3 points above the total effect, e.g. expected scores of 444), followed by Dorchester Two (2 points above the total effect), and finally Dorchester Four (6 points below the total effect). With a mean of 440

and a standard deviation of 39, this produces a margin of error of 2.866 points, e.g. scores between 437 and 443 are indistinguishable from the average performance. This means that the average performance we expect to observe in Dorchester Four would be significantly different than the expected performance overall. From this we may conclude that the selection bias in Dorchester Two significantly affected the performance of students in that district resulting in wide variation and therefore a lack of statistical significance for Dorchester Four's scores.

Analytical Methods

Because the goal of this analysis was to establish the degree to which we can say there is an effect from UFLI and what that effect is, it is necessary to control for confounding covariates. Unfortunately, observational data faces a significant challenge in that the assignment mechanisms to test and control groups are often ambiguous and can create confounding covariates when the selection mechanisms are not known a priori by a researcher. The goal of matching data is to prune the data set so that the remaining data sets are balanced between the treatment and control groups, e.g. simulating a true random experiment. This analysis uses two methods of analysis, one which uses traditional modeling techniques, i.e. multilevel modeling, and another that uses coarsened exact matching methods (Iacus, King, and Porro 2011), to simulate experimental design. The CEM method is used to predict individual test scores which are then sorted based on the a priori categories of No Met, Approaches, Meets and Exceed that are provided by the SC Department of Education. Multilevel modeling is also used to predict individual test scores and the results are compared to the CEM methods. Based on the predicted test scores, the individual students are sorted into

groups of (1) Meets/Exceeds or (0) Approaches/Not Met. Multilevel logistic regression is then used to determine the probability that a student scores Meets/Exceeds based on specific criteria including whether the student was in UFLI, prior ability, parental education level, poverty, race and gender. Grouping variables are student class and school.

To estimate SC READY Scores, we use the average grade-level coefficients for DRA-2 effects on SC READY and MAP coefficients on SC READY. In other words, for every 1-point increase in second grade Spring DRA-2 score, there is a corresponding increase in SC READY of X . We then use the baseline (no effects from UFLI) DRA-2 score to estimate the Baseline SC READY score. We then add in effects of UFLI to the DRA-2 score and estimate a second SC READY score that we can compare baseline to effect score. This is like estimating a production function in economic analysis. The model is robust to changes in poverty levels and parental education levels as well as the effects of race and gender.