Trident United Way

Summary of the Trident United Way Reading by Third Project

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

Teacher satisfaction varied between schools and over time. In year one, all 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 what began at a level A-1 (non-reader) and in three months was at a DRA level 10 (mid-first grade capability).

Teachers in year oner 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. This is a pattern which emerged each year, i.e. teachers were initially reluctant but once they began to see results, they bought into the program.

Fidelity of Imlementation

To assess fidelity of implementation, teachers recorded themselves delivering a UFLI lesson, typically recording with 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.

Several schools showed significant difficulty with UFLI and, according to the Lastinger staff, did not implement with fidelity, at least initially. Newington Elementary and Flowertown Elementary did not display initial buy-in to the program. In fact, when teachers were anonymously polled regarding what improvements the school could make for the following (2019-20) academic year, one response was, “get rid of UFLI.” Teachers in Newington and Flowertown were initially hostile to the program, but as with other schools they eventually bought once they saw the results. A common theme in Newington and Flowertown is that they were not the schools that were initially selected for years two and three of the project. Rather, Dorchester Two changed the schools receiving the training. This was concerning since both schools had new principals, rather than the experienced principals required by the grant, although both were Title I and both met the performance criterial for the grant. The change did provide valuable insights on the role of the principal in the UFLI program, e.g. a new principal who is just establishing their school culture will likely struggle given the time and scheduling requirements necessary for successful UFLI implementation. This bore itself out in the average performance within these schools and may have been a contributing factor for the poor initial teacher buy-in.

Challenges to the Evaluation

Challenges existed from a data collection and an implementation standpoint. All districts faced challenges in finding time to implement the one-on-one sessions. Although the

districts shared best practices during quarterly meetings and at the individual UFLI trainings, the ability to implement the scheduling suggestions was ultimately up to the school principal. A general refrain was year-over-year from district representatives was that while the teachers found it difficult to fit the sessions into their schedules, they nonetheless were eager to engage in the UFLI sessions. When asked about this, the representatives responded that the teachers felt that they were empowered with regards to how to teach their students and that the success they saw in their students was motivating for them.

Challenges in data collection included willingness or ability of the districts to provide socioeconomic data on students, i.e. WIC, TANF, Medicaid status. Given the known effects of poverty on educational outcomes this makes it difficult to provide accurate feedback as to the success of the program since we cannot control for socioeconomic priors that are known to affect educational outcomes. To address this, the author reached out to the South Carolina Department of Education about providing student level poverty data.

A second challenge for data collection was in the consistency of the tests used for assessments. State testing requirements left the districts with multiple options, so coordination was key to collecting measures that were comparable across all districts. The districts initially stated that they would all conduct the DRA-2 assessments along with the TOWRE-2 assessment. The South Carolina State Department of Education changed requirements from DRA-2 to KRA as the kindergarten assessment prior to the beginning of the 2017-18 academic year, leaving districts with a choice of which test to administer.

The lack of consistency in the exams does not hinder the ability to evaluate changes in student reading since all districts administered exams aimed at measuring the ability to read informational text. The lack of consistency does hinder the ability to determine the depth of student learning since informational text scores do not provide information on automaticity in student reading, i.e. whether reading was an automated response or whether students were simply able to read, but with increased levels of effort.

A third challenge to the evaluation was staff turnover within the districts. In each district, the individual responsible for the collection and dissemination of the data to TUW turned over at least once. Therefore, the data collection practices, data sharing agreements, and data distribution had to be retaught or reviewed again on multiple occasions.

Perhaps the most daunting challenge to the evaluation came in the form of COVID-19 and the state's response. The initial intent of the project was to assess improvements to third grade reading. For AY 2019-20 though, the state received a waiver from federal testing requirements meaning that TUW did not have access to the data necessary to show an effect from UFLI. Given this, TUW estimated the effects of UFLI using the known the coefficients for DRA-2 on SC READY. The evaluator was therefore able to estimate what the performance of UFLI students was most likely to have been after controlling for race, poverty status, parental education levels and between class and between school differences. This was not possible for CCSD though. This was because CCSD focused on Pre-K which does not have a mandated federal exam. The exam that is used by CCSD (MyIGDIs) was not given at the end of the year

given the school closures. Therefore, there is not data for CCSD's performance in year three.

General Findings

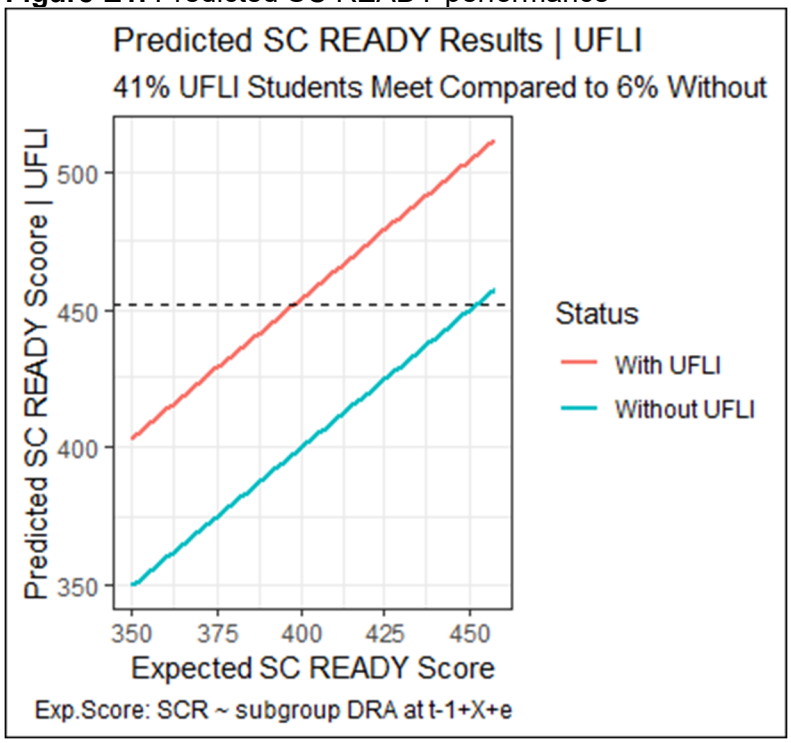
Over a three-year period a total of 4,775 students were taught by a UFLI trained teacher. This is comprised of 1,054 (Berkeley) 1,515 (Dorchester Two), 1,619, and 567 (Charleston). Of the 4,775 students taught, 729 students were struggling readers targeted for UFLI. The selection of students into UFLI 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 as 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 form any other UFL 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 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 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.