STRATEGIC DATA PROJECT

SDP FELLOWSHIP CAPSTONE REPORT

Early Elementary On-Track Indicators Leading to Third-Grade Reading Proficiency

Jing Che, Rochester City School District

Patty Malgieri, Rochester City School District

Vicky Ramos, Rochester City School District

Hannah Page, Office of the State Superintendent of Education

Anna Holt, Tulsa Public Schools

SDP Cohort 5 Fellows

Strategic Data Project (SDP) Fellowship Capstone Reports

SDP Fellows compose capstone reports to reflect the work that they led in their education agencies during the two-year program. The reports demonstrate both the impact fellows make and the role of SDP in supporting their growth as data strategists. Additionally, they provide recommendations to their host agency and will serve as guides to other agencies, future fellows, and researchers seeking to do similar work. *The views or opinions expressed in this report are those of the authors and do not necessarily reflect the views or position of the Center for Education Policy Research at Harvard University.*

Framing the Problem

The No Child Left Behind legislation mandates the establishment of K–12 data systems in state and local educational agencies across the nation. While this effort to capture data should be applauded, the use of data analytics to inform decision-making lags behind (Daly, 2013). This issue has generated concerns, particularly in early childhood education and in the primary grades of struggling urban districts. Challenging questions such as how the quality of early childhood programs should be defined and measured and how to track students' cognitive and social-emotional growth throughout the primary grades have been raised in both research and practice. Though the effectiveness of early indicators for high school graduation have been frequently examined, field-tested, and utilized in educational agencies across the country (Davis, Herzog, & Legters, 2013), the indicators for primary grades in urban settings remain elusive, except for a few efforts (e.g. Massachusetts Early Warning Indicator Systems).

Unlike students in upper grades, students in Pre-Kindergarten (Pre-K) to third grade move through different developmental phases very quickly. Therefore, investigation of the comprehensiveness of these early success indicators in multiple cognitive and non-cognitive domains seems critical. Furthermore, the well-known 30 million word gap between students from high-income families and those from families on social welfare (Hart & Risley, 2003) provides a sense of urgency to study this issue and to identify on-track indicators as early as possible and align them with the appropriate targeted supports. This capstone project explores the full array of early indicators for third-grade reading proficiency using a multi-site case study approach. Early indicators are researched and field-tested in Rochester, NY, Washington DC, and Tulsa, OK to examine the feasibility of identification and implementation of Early Warning Indicator Systems (EWIS) in three diverse urban districts. Overall, we intend to address the following policy questions: 1) What are the Pre-K to second-grade key early indicators that predict third-grade success? and 2) How do we determine and apply indicators that track developmental and social emotional growth?

Literature Review

Recent headlines about standardized scores in early grades have sparked a national debate about the developmental appropriateness of standardized testing for grades as early as first grade (Sapers, 2014). Research has demonstrated the relationship between student characteristics, graduation rates, and even earnings potential (Pascarella & Terenzini, 2005). Consequently, policymakers continually seek strategies to better support school populations who are at risk of not graduating with interventions to ensure positive outcomes. As it becomes increasingly difficult to steer students towards graduation after they enter high school, practitioner as well as researchers have started to look towards interventions in earlier grades (Mac Iver, 2009). Hernandez (2011) found that third-grade students who do not read proficiently are four times more likely not to graduate from high school on time than third-grade students who read proficiently. The oft-cited Abecedarian Project and the Perry Preschool project have shown that benefits from high-quality early childhood environments last into adulthood (Campbell, 2001). A 1997 study tracked the growth of a first-grade cohort in Baltimore for 14 years and found that early elementary characteristics influence high school dropout rates independent of socio-demographic factors (Alexander, 1997). Additional research indicates that 50% of eventual dropouts can be identified as early as sixth grade by examining attendance and discipline trends along with course performance (Iver, 2009). In sum, early literacy indicators seem to have a strong relationship to later success in school.

Nonetheless, despite the Obama administration emphasizing early childhood education, the reality is that 37% of U.S. fourth graders fail to achieve basic levels of reading achievement according to the National Assessment of Educational Progress (NAEP) (Lonigan & Shanahan, 2009). Consequently, identifying students who need intervention earlier in primary grades has become the first step to stop the leaking of the pipeline.

Existing On-Track Indicators

Early warning systems (EWS) aggregate information from diverse data systems with the purpose of providing useful information earlier to education practitioners. Massachusetts developed an Early Warning Indicator Index for rising ninth-grade students that provides an estimated probability of being at risk (AIR, 2013). After developing this tool, school leaders and

teachers asked that this index be extended to first grade and suggested using third-grade proficiency on the Massachusetts Comprehensive Assessment System (MCAS) as the targeted outcome. Selecting outcomes that are closer to the time of the indicator mitigates the possibility of teachers and school leaders labeling early elementary children as potential high school drop-outs and allows for the targeted development of interventions leading to third-grade proficiency (IES, 2015).

The Balfanz Attendance Behavior Coursework (ABC) Model

The use of attendance, behavior and coursework metrics—commonly referred to as the ABC model—is giving way to more sophisticated models that use advanced statistical methods beyond linear regression (Knowles, 2014; Balfanz, 2000). Attendance trends are often indicative of later academic success and have been shown to impact educational outcomes as early as Pre-K (Chicago, 2014). The Massachusetts State Department of Education measures attendance by calculating the rate of days attended over total days enrolled (AIR, 2013). Similarly, Montgomery County Public Schools examined an array of third-grade academic and behavioral indicators for students' sixth-grade performance, and found that absences in the first-semester of third grade are significantly and negatively associated with academic outcomes in sixth grade (West, 2013).

Behavioral measures can provide useful insights for early elementary EWSs. Rates of suspensions and expulsions have been used as predictors of academic success, as well as for the development of behavioral intervention plans (Balfanz, 2000). Taking a prevention-focused approach as soon as challenging behaviors are presented has been shown to be more effective than a reactive approach later in a student's educational career by reducing both missed instruction time and the stigmatization surrounding many disciplinary consequences (Brazelle, 2015).

Poor performance in early grade subjects has been included in EWSs as an indication of potential academic challenges (Balfanz, 2000). While course completion is measured differently for early elementary grades, performance in grades as early as first grade has been directly linked to performance in middle and high school grades (MacIver, 2013).

Additional Indicators

While attendance, behavior, and coursework indicators form the backbone of most EWSs, other indicators may also be appropriate to include in the model. Demographic characteristics of the students are generally included in early warning models and most commonly include race, ethnicity, parental education levels, and economic disadvantage; they are often used as control variables to test the impact of variables that can be adjusted with behavioral or policy shifts. For example, the Early Childhood Longitudinal Study-Kindergarten Cohort followed more than 14,000 children beginning in 1998 and investigated the relationship between individual characteristics and achievement outcomes. The study found that a mother's level of education is a significant predictor of Kindergarten readiness (Tang, 2014). The role of trauma and disruption within the family context can also be included in these models, as these can create additional stress on students and impact their academic success; further, studies have found that this impact can be greater in the early years of childhood (Alexander, 1997).

Measures of students' social-emotional well-being are increasingly being considered along with academic outcomes. The Los Angeles Unified School District includes the development of non-cognitive skills in its School Quality Improvement Index. Similarly, regulation and persistence in early grades are now included in many other school systems' visions. Third-grade students who return homework without teacher prompts have the largest positive associations with sixth-grade academic outcomes in a study of Montgomery County Public Schools (West, 2013). Academic persistence and self-regulation in early grades are also important indicators for school success. Studies have linked negative social emotional development with school readiness in children under the age of 6 (Shonkoff, 2000).

Table 1. Table of Frequent Indicators in Early Warning Systems

Indicator	*Common Sub Indicators*
Demographic	Socioeconomic Status, Gender, Race, Family Type, Over-age
School Participation	Attendance, Retention, Mobility, Special Education
Historic	Previous Program Participation, Student Mobility

Measurement and Validation

Measuring growth in any of these areas becomes increasingly complicated in the earlier childhood space, where proficiency and growth are calculated on a developmental trajectory unique to the individual child, and measured in developmental phases rather than grade specific milestones of mastery. While there is increasing agreement among policymakers about the importance of early childhood education at the federal, state, and local levels, the alignment of these varied stakeholders brings with them unique challenges.

Current Models & the Strategic Data Project Capstone Group

Massachusetts currently has the only model that incorporates first-grade data, although other states are rapidly developing pilots to enter this space.

Table 2. SEA/LEA Early Warning Systems Examples

SEA/LEA	*Year Launched*	*Earliest Grade*	*Risk Levels*	*Indicators included*
Massachusetts	2011	1st	3 (Low, Moderate, High)	Attendance, school move, suspensions, low income, special education, English-language learner (ELL), gender, urban, over-age, Title I (school and targeted)
Pennsylvania (Philadelphia)	2008	Middle School		
Virginia	2009	9th	Index Score	Attendance, course failure, GPA, discipline
Montgomery County, MD	2013	1st	3 (Low, Moderate, High)	Attendance, course failure, GPA, discipline
Chicago (CCSR)	2012	Pre-K	NA	Primarily investigated impact of categorical attendance patterns
Portland	2008	9th		
Baltimore		6th		Chronic absence, course failure, over-age, multiple suspensions

This paper presents analyses from three comparable districts focusing on a different age span within the Pre-K3 to third grade continuum. All three case studies attempt to identify indicators for early literacy and numeracy. While a full range of demographic controls are

included in the analyses, all three cases focus on finding school-based, action-oriented levers to develop targeted interventions for improving student outcomes.

Table 3. Capstone Districts and Grade Spans Analyzed

District	*Case Study Summary*	*Pre-K3*	*Pre-K4*	*K*	*1st*	*2nd*	*3rd*
Rochester	Impact of student and program-level predictors between K and third grade on third-grade reading						
Tulsa	Impact of second-grade data on third-grade reading						
Washington, DC	Using Pre-K to third grade data to predict and support Kindergarten readiness						

Case Studies

Rochester Case Study—Indicators for "Reading by Third Grade"

Context. Rochester City School District is a mid-sized urban district serving approximately 30,000 students. Once a city thriving with industries, Rochester is now facing grim realities for the education of its students. Rochester, with the majority of its student body being students of color, has the lowest test performance and highest poverty rate (83%) among the Big Five urban districts in New York State. With only 5.5% of the third graders in the district reading at grade level, as measured by the year-end New York State English Language Arts (ELA) test, district leadership sensed an urgency to improve early literacy and to stop the leaking of the pipeline early on. This capstone project, conducted by Jing Che, Patty Malgieri, and Vicky Ramos, (SDP Fellows), aligns well with District's Action Plan and its Academic Priority of "Reading by Third Grade". This project analyzes the multitude of longitudinal administrative data—ranging from students' demographic, attendance, behavioral, academic, and

developmental records in Pre-K to third grade—to provide district leaders and school administrators with actionable recommendations.

Demographics of the third-grade cohort (*N*=2183) represent a typical student body in a high-poverty and underperforming urban district with 92.6% of students receiving free and reduced-price lunch and majority of students being African American and/or Latino. Furthermore, students receiving special education services/students with disabilities (SWD) constituted 17% of the cohort group, while the proportion of students with Limited English Proficiency (LEP) was 12%. Sixty-one percent of the students in this sample were once enrolled in a Universal Pre-K (UPK) program in the district, and one out of ten students were retained a grade level before reaching third grade (See Table 4).

Table 4. Third-Grade Student Demographics (Rochester)

Demographic Indicator	Percentage in 2013–2014 Third-Grade Cohort
Free/Reduced-Price Lunch	92.6%
Underrepresented Minority	87.7%
SWD	16.9%
LEP	12.0%
Pre-K Participation	61.3%
Ever Retained a Grade	10.4%

Policy/Research Question. Fellows were tasked with designing and rolling out an early indicator system for Third-Grade Reading Proficiency for Rochester City Schools. Therefore, the policy questions central to this capstone work are: What are the Pre-K to third-grade key indicators that predict third-grade reading proficiency? How are students faring across these early grades?

Project Scope and Timeline. This project marked the district's first comprehensive effort to longitudinally track student outcomes in the early grades, making it a strong case study

for the trajectory from rich data to rich information at a district level. Prior to this project, most of the data reporting served accountability purposes, and little work had been done to connect student data from different domains and databases to present a complete student profile. In the beginning phase of this project, an extensive literature review was conducted to identify potential indicators, and all available data elements in the district's databases were compiled for the analysis. Consequently, key stakeholders, including district's executive cabinet members and directors of early childhood education, were engaged in the discussion. The leadership decided to roll this out as a pilot with only two elementary schools in the first year, so principals at these two schools were also kept abreast throughout the process. A districtwide implementation of a data dashboard that includes an early indicator system for third-grade reading proficiency will follow the pilot and is scheduled for summer 2016.

Methods. Lasso Regression was used as the primary analytic approach to identify the best predictors. Lasso is a shrinkage and selection method that balances model fit and parsimony (Tibshirani, 1996). Students' school attendance, behavioral records, and academic performance scores from Pre-K to second grade were regressed against a host of student demographic characteristics and other control variables to test reading proficiency outcomes[1]. Meanwhile, school effects were entered as fixed effects in the model. Consequently, Ordinary Least Squares (OLS) regression analysis using the selected key indicators from the Lasso method was performed to estimate the predicted values of the reading scores. This predicted value, representing a weighted combination of all the indicators, was denoted as the Composite Success Factor in this analysis. Next, two competing methods for determining cut scores for the above key indicators were compared. One is called Receiver Operating Characteristic (ROC) Curve analysis (Bowers, Sprott, & Taff, 2013), which helped identify different possible cut scores for the composite success factor and key actionable predictors based on the trade-off between the accuracy of

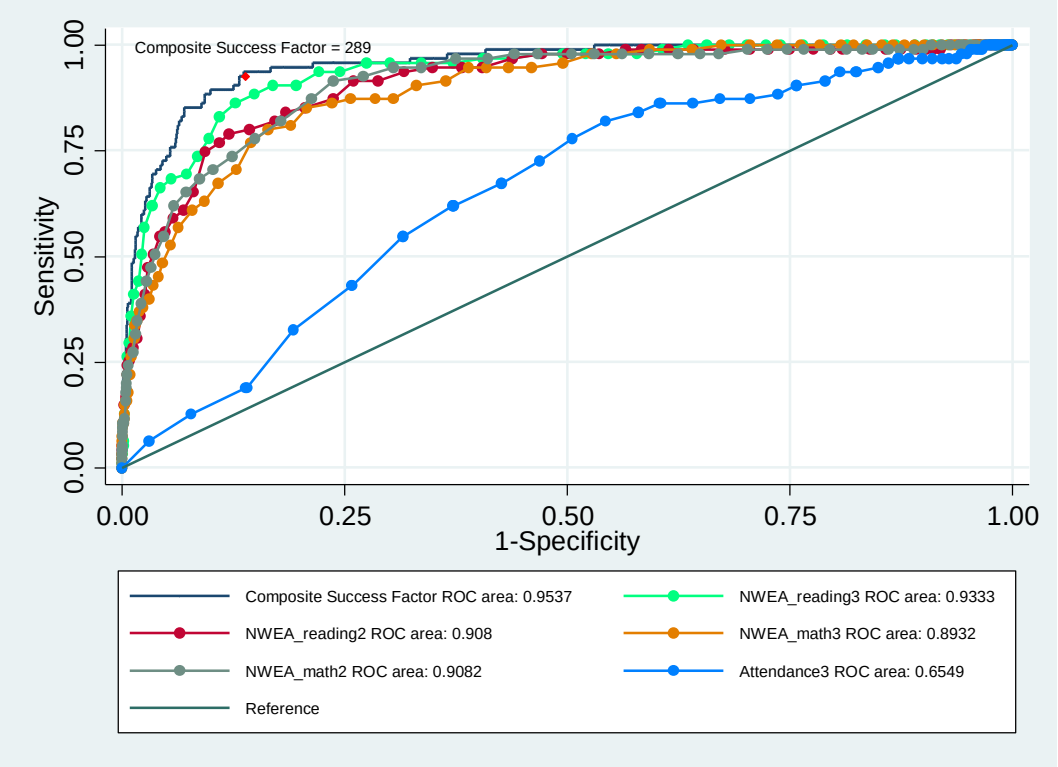
[1] Variables entered into the Lasso model included: student demographic characteristics such as race/ethnicity, gender, poverty status, Limited English Proficiency status, and Students with Disabilities status, students' attendance and behavioral measures such as school-year attendance rates and suspension records in Grades K–3, and whether or not students have ever been retained, and students' academic performance measures such as whether or not students have ever had enrolled in Pre-K programs, Kindergarten screening results, and district's end-of-year benchmark tests (i.e. NWEA MAP Reading and Math tests) in Grades K-2.

prediction and sensitivity of the differentiation power of these metrics. The other is a Classification and Regression Tree (CART) method, which is a nonparametric approach for classifying students as proficient or non-proficient in a "tree" format based on a set of if-then statements (Koon & Petscher, 2015). The CART method only takes into account the key predictors, keeping the ones that are most significant, and empirically grows a "tree" to identify a combination of cut scores for these indicators. Finally, relationships between Pre-K enrollment and school attendance and academic performance in grades K–3 were also explored to evaluate the potential influence of Pre-K on later school success in a separate OLS regression analysis.

Results. The key actionable predictors for third-grade reading proficiency identified by this study include district's benchmark tests, Northwest Evaluation Association's Measures of Academic Progress (NWEA MAP) Reading and Math, in first and second grades, and second-grade attendance. See Table 2 in Appendix for details. Meanwhile, Figure 1 shows the results of ROC curve analysis and the prediction sensitivity and error of different cut scores on these key indicators[2]. The x-axis represents the false positive rate (1-Specificity), the percentage of students who are predicted to pass but fail out of all who actually fail third-grade ELA.. The y-axis is the true-positive rate—the percentage of students who actually pass the ELA tests out of those identified as passing. There is a clear trade-off between error in prediction and sensitivity of the model to pick up the right candidates. In this specific case, we think we can tolerate a higher level of prediction error, even if we falsely identify some students as candidates for passing the ELA tests, who still fail. The rationale behind this decision is that we want to cast a wide net to create a potential candidate pool for early intervention as long as the net is not too wide for us to lose capacity. Since we do not have any prior academic baseline for these third graders, historically, we were not able to correctly identify them in second grade for early intervention. To this end, the ROC curve analysis empirically determines cut scores and is a big step forward for the district's data use.

[2] The cut scores based on the ROC curve analysis make false negatives (Type II error) more costly than false positives (Type I error), i.e. giving more weight for sensitivity than 1-specificity.

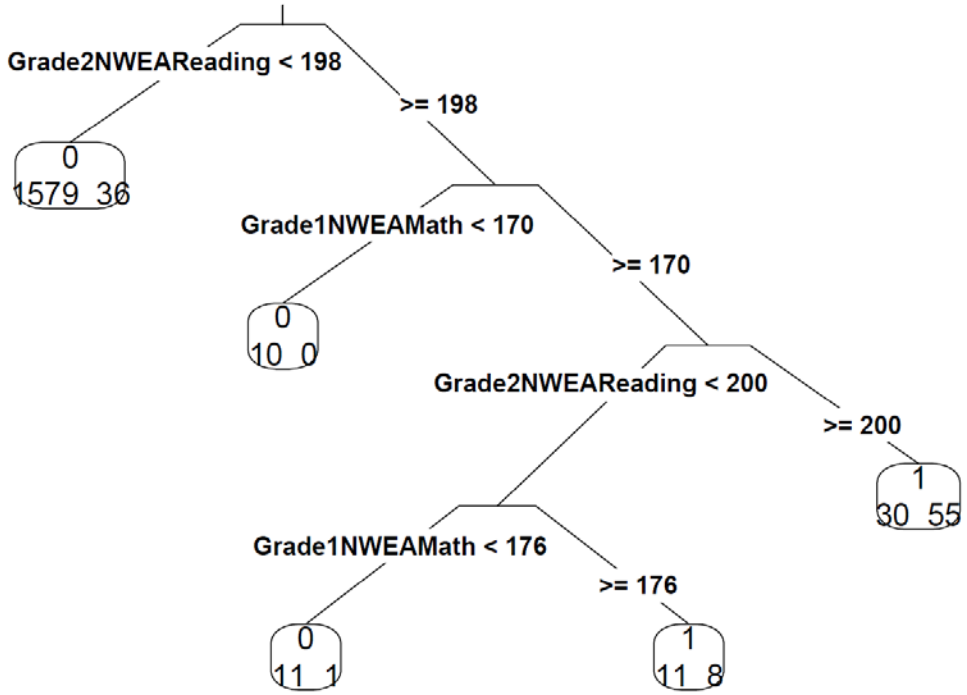
Figure 1. ROC Curve Analysis of the Composite Success Factor and Key Actionable Indicators (Rochester)



As a cross-validation, an alternative method—the Classification and Regression Tree—is also used to determine the cut scores empirically. Under this method, the district's benchmark tests (NWEA MAP Reading and Math) in first and second grades, second-grade attendance, and SWD status were identified as key indicators. Each indicator was entered to grow a "tree," but only second-grade MAP reading and first-grade MAP math were kept due to considerations of parsimony. Again, false negatives (failure to identify those who actually passed the exam) were weighted three times more costly than the false positives (failure to identify those who predicted to pass but who actually failed the exams). The cut scores for proficiency were as follows: (1) scoring greater than 200 in second-grade MAP reading, (2) greater than 198 in second-grade MAP reading and (3) greater than 176 in first-grade MAP Math. This corresponds to a 98% sensitivity and 61% false positive rate on the ROC curve chart, with only 104 students who met these criteria in this cohort. Different weightings for the type I and type II errors do

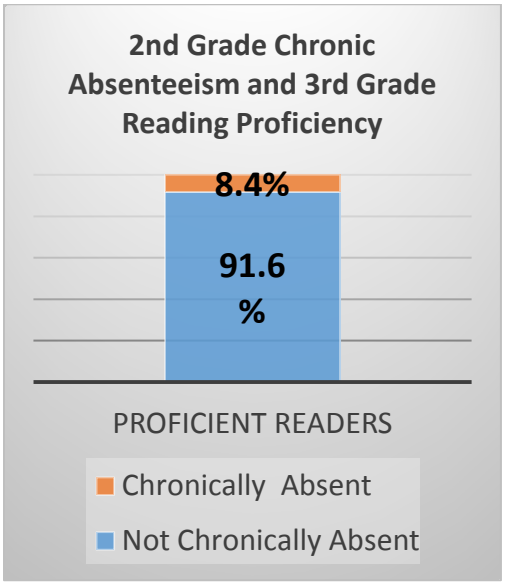
not seem to exert much difference. Such a cut score for second-grade reading is more than 2 standard deviations above the national average. Therefore, even though the CART method shows great advantage in visualizing the cut scores, it does not seem to provide any practical support for our administrators in this particular analysis, likely due to the lack of variance in the outcome measure, with only 5.5% proficient readers. It might still be appealing to practitioners in some of our future projects.

Figure 2. Results of Cut Scores for Key Indicators from the CART Method (Rochester)



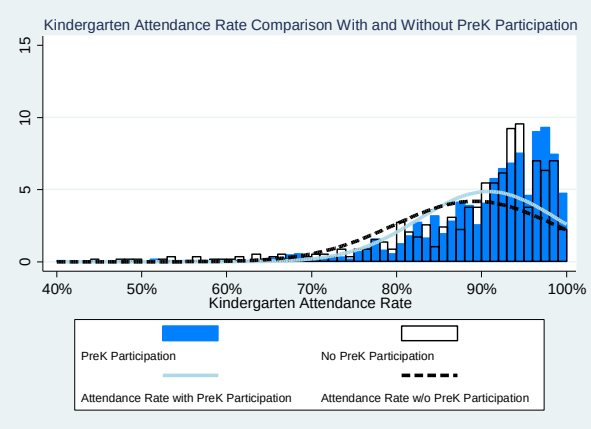
Furthermore, Table 4 and Figure 3 reveal that second-grade attendance is marginally significant for third-grade reading proficiency. The vast majority (91.6%) of the proficient readers at third grade had attendance above 90% in second grade, which means that an overwhelming majority of the students in the examined cohort had no problem with chronic absenteeism.

Figure 3. Second-Grade Chronic Absenteeism for Proficient Third-Grade Readers (Rochester)

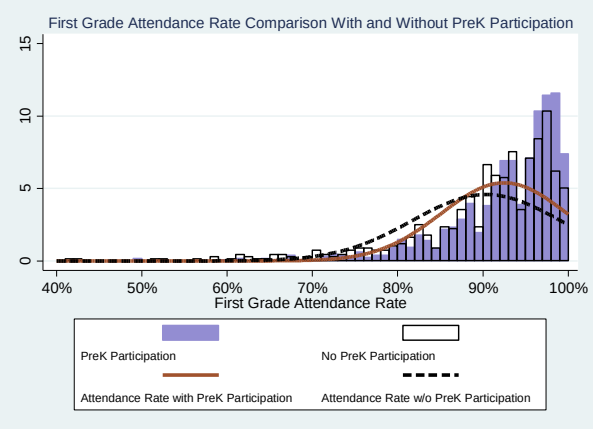


Although Pre-K participation does not seem to have a significant impact on students' third-grade reading proficiency scores, students in the 2013–14 third-grade cohort who participated in Universal Pre-K programs, be it half-day or full-day, were more socialized into the classroom routines and were more ready to engage in learning. Pre-K participation has a long-lasting effect on attendance in grades K–3 and academic performance measured by NWEA MAP tests in grades K-2 when regressed against a host of demographic controls (See Figures 4a–4d). These students have, on average, almost 2% higher attendance rate than those without UPK in all early grades from K to 3, after controlling for student demographic characteristics. Furthermore, students with UPK also had significantly higher beginning Kindergarten screening scores measured by the Child Observation Record in the fall of their Kindergarten year (On average, 0.2 standard deviations better, i.e., almost a quarter ahead of their non-UPK peers), and significantly higher end-of-year Kindergarten readiness scores in the spring. In addition, students with Pre-K participation performed significantly better than their counterparts in first- and second-grade NWEA MAP reading tests—the difference is approximately .15 standard deviation units.

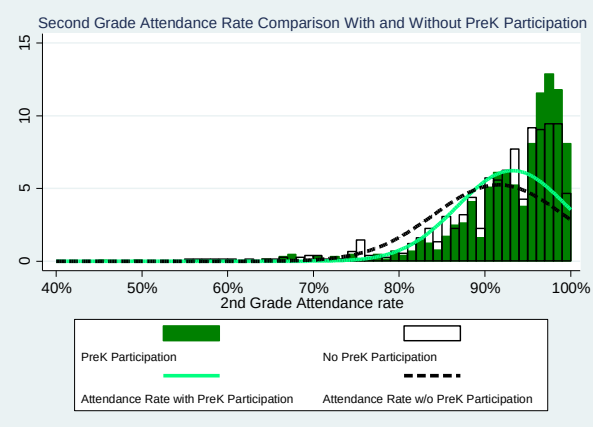
Figure 4. Comparison of School Attendance between Pre-K Participation versus No Pre-K Participation



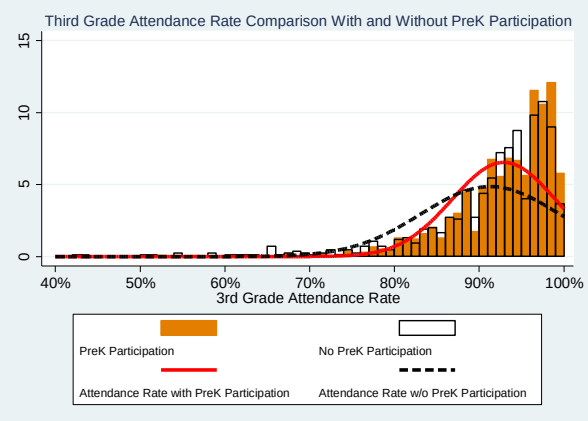
4a. Kindergarten Attendance



4b. First-Grade Attendance



4c. Second-Grade Attendance



4d. Third-Grade Attendance

Consistent with most of the previous literature, some of the demographic factors including race/ethnicity, poverty, disability status, and Limited English proficiency status still exerted prominent influences on our students (See Table 2 in Appendix). Additionally, significant school contextual effects were observed. Certain schools performed better than others overall.

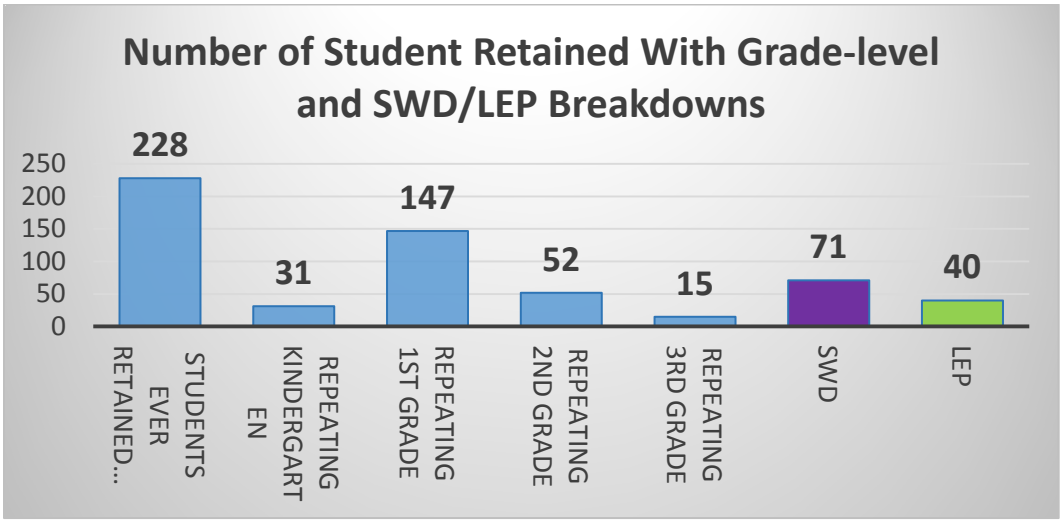
There were mixed findings regarding grade retention. Retention is not significantly related to third-grade reading scores after controlling for other factors. However, students

ever-retained (228 in total) have significantly lower attendance rate in K–3 and lower first- and second- grade NWEA reading scores than their never-retained peers. As shown in Figure 5, 31% of the students who have ever been retained are students with disabilities, and 15% are students with Limited English Proficiency, both of which are disproportionally higher than the student composition of the entire cohort. Moreover, the Academic Standards and Policy (ASAP) in the district has restrictions on elementary and middle school age children. The business rules include:

- No student is allowed to start the school year in a K–6 school if he/she will turn 14 years of age during that school year.
- No student is allowed to start the school year in a K–8 school if he/she will turn 16 years of age during that school year.

These rules suggest that students are only allowed to be retained once in grades K–6. However, as shown in Figure 5, 17 students in this cohort have repeated grades twice before the end of third grade, which might require further investigation.

Figure 5. Breakdowns of Students Ever Retained by the End of Third Grade



Discussion. This project is one of the district's first endeavors to examine student achievement and growth data in early grades in a longitudinal and comprehensive fashion. To address the research questions we posed earlier, fellows have the following recommendations:

Research Question 1: *What are the Pre-K to third-grade key indicators that predict third-grade reading proficiency?* The key actionable early indicators identified by this study include successful passing of the district's benchmark tests (NWEA MAP) in reading and math during first and second grades, as well as second-grade attendance (90% attendance cut off). By differentiating students into groups based on benchmark test results at different checkpoints throughout the early grades, educators can identify a candidate pool for targeted interventions. The ROC curve analysis identifies the bucket of students with potential for Level 3 or above. The analysis can also show the cut scores for identifying potential Level 1 (bottom level for proficiency) and level 4 (highest level for proficiency) students to provide appropriate supports. Furthermore, although second-grade attendance is not as effective as MAP tests in predicting reading proficiency in the Rochester case, that attendance matters sends a strong message to schools, parents, and communities about the importance of keeping students in school. Chronic absenteeism at this age is still largely attributable to parental circumstances. The findings should generate meaningful discussions within our community to seek possible solutions regarding potential conflicts between parent work schedule, sibling care, and student transportation. Superintendent Dr. Vargas is leading the campaign to achieve 95% attendance in our district; bus passes were distributed to parents of students in Pre-K and primary grades.

Research Question 2: *How are students faring across these early grades?* Although universal Pre-K programs and grade retention do not have a significant relationship with the third-grade reading outcomes, they are both significantly associated with some of the more intermediate outcomes across the early grades. Students with UPK are faring better than those with no UPK across the early grades in terms of academic behavior and school readiness. However, students who have repeated grades do not seem to fare as well. They were behind early on and did not seem to catch up later in third grade. Additionally, disproportionally higher numbers of students with disabilities and limited English proficiency were retained and also not shown progress later on. Multiple retentions in early grades, which is against our ASAP policy,

also requires attention. Overall, this calls for immediate action to revisit and reexamine the retention data and policy.

Key Takeaways.[3] First, absent baseline state exam data available prior to third grade, this analysis identifies potential proficient readers for early intervention by the end of second grade. In other words, the district can cast a wide net for early intervention. Second, the use of an on-track, off-track cut score for first- and second-grade NWEA reading/math benchmark tests and the related composite success factor is likely to be a more effective and friendly way to communicate with practitioners than the actual results of the predictive models. Third, significant and positive association of Pre-K participation with attendance and academic outcomes serve as strong evidence for the district's implementation of full-day Pre-K program and summer Pre-K programs. One theory emerged from the field is that Pre-K is about socializing students into the classroom community, setting the norms, knowing the expectations and routines during the school day and not just about the academic outcomes at this age. Fourth, benefits of consistent use of benchmark tests are demonstrated in this analysis. The district has had a history of frequently changing benchmark tests due to curriculum or leadership change, which has made vertical test comparison for student growth difficult. Last, retention in early grades should be re-visited with further data cleanup and analysis.

Next Steps. As this capstone project evolved, the team has come up with a few action items. First, we will test the current model using new waves of third-grade cohort data and continue to track the 2013–14 third graders into their fourth grade. Meanwhile, we will field-test the models with administrators and teachers at two elementary schools in the district. The two schools selected for the pilot sit on two extremes of a spectrum in terms of school climate and performance. Although both principals are strong advocates for data use in decision-

[3] This study also has several limitations. First, even though the data naturally presents itself in a nested structure, mixed-effect models were not used in this analysis due to the fact that there has yet been a Stata package marrying Lasso method with mixed-effect models. Second, the collinearity issues among the multitude of indicators included in the model continue to create confusion over the predictive power of each individual measure. Third, the predictive power for future third-grade cohort using New York State exams is reduced due to the mass opt-outs this past year. All these limitations add qualifications to the interpretation and inference of the findings from this cases study.

making, school cultures are drastically different. One school has the highest student proficiency rates in ELA and Math in the district with a strong teacher-led data team that examines student assessment data quarterly for goal-setting and intervention strategies. The other school is state-cited as a persistently failing school with enormous challenges that include: student discipline issues, lack of quality instruction, and low staff morale. In collaboration with the two schools, the fellows have sought feedback and comments from practitioners regarding appropriate cut scores for key indicators, and thus created more effective and user-friendly ways to use benchmark test (NWEA-MAP) data. Based on some initial conversation with principals, school-based benchmark tests might add some richness to the picture. However, the burning questions from the school administrators' perspective revolve around the action items: *how much growth is needed to close the achievement gap? What timeframe would look realistic? Should it be a one-year, two-year, or even three-year plan for different students? And how should intermediate goals for students be set?* School administrators would like to take both growth and proficiency into consideration when differentiating students for appropriate interventions, which requires some innovative and interactive ways of displaying data that are used in daily school planning. Fellows will explore these issues moving forward with the ultimate goal of creating a district-wide comprehensive dashboard application for the Early Indicator System.

Second, we desire to reexamine the district's retention policy and the impact of retention to make strategic recommendations for changing policy and practice. Third, district leadership will want to share the findings about the impact of Pre-K attendance with the community to ensure implementation of a full-day Universal Pre-K districtwide. Finally, a deep-dive analysis of current Kindergarteners with varying dosages of Pre-K intervention (i.e., no Pre-K, half-day Pre-K, full-day Pre-K, Pre-K with/without summer programming) would also add richness to the findings of this study, and would encourage further policy discussions around implementing UPK in the district.

Tulsa Public Schools—Indicators for Third-Grade Reading Proficiency

Context. Tulsa Public Schools (TPS) is a mid-sized urban district serving approximately 40,000 students. In 2014, 47% of all third-grade students were reading on grade level on the

Oklahoma Core Curriculum Test (OCCT). This composite number masks wider disparities by economic-disadvantage status and ethnicity. Tulsa Public Schools had an 82% overall rate of students in this third-grade cohort receiving free and reduce-priced lunch. Additional demographic information for the 2013–14 third-grade cohort is provided in the table below.

Table 5. Demographic Profile of the Tulsa Public Schools 2013–14 Study Cohort (Tulsa)

Demographic Indicator	Percentage in 2013–14 Study Cohort
Economically Disadvantaged (FRL)	81.8%
Students with Disabilities (Special Ed Program participation)	24.7%
English Language Learners	12.9%
Traditionally Under-Represented Minority	70.6%

Importantly, district-wide, 41% of economically-disadvantaged students scored on grade level on the 2013–14 vs. 69% of non-economically-disadvantaged students. Additionally, 62% of White/Caucasian students were on grade level vs. 33% Black/African American and 38% Hispanic/Latino students.

The state of Oklahoma recently passed the Reading Sufficiency Act. While this Act is comprehensive and requires school actions at the K–3 level—including the development of reading sufficiency plans for each school site, the use of differentiated reading interventions, and access to summer reading programs for low-achievers—the policy is most notable for its third-grade retention requirement. Beginning in 2013–14, students scoring unsatisfactory on the OCCT face mandatory retention unless they qualify for one of six good-cause exemptions. An initial 35% of students in Tulsa Public Schools were at risk for retention in 2013–14 because of an Unsatisfactory OCCT score, and 18% were ultimately retained after good-cause exemptions.

The district needed to validate their choice of a third-grade alternative benchmark assessment as an accurate predictor of final student retention and scaled student score outcomes. The validation and predictive modeling comprised Phase I of the capstone work. At

the same time, the district sought to use the benchmark nature of these predictors to track gains at multiple time points prior to OCCT administration and to provide these tools to our K–2 teachers to support monitoring of classroom progress.

Table 6. Results of Regression Analysis (Tulsa)

Indicators (Retained in Final Model)	β	*s.e.*	*t*	*P*
Intercept	105.65	15.05	-7.02	0.000
Black or African American	-18.62	3.31	-5.62	0.000
Hispanic or Latino	-3.24	5.07	-0.64	0.522
Native American or Indian	-7.37	4.96	-1.49	0.137
Multi-Ethnic/2 or More Races	-7.69	4.12	-1.86	0.062
Asian/Pacific Islander	-0.07	10.49	-0.01	0.995
(Reference Category: White)				
Free/Reduced Lunch	-11.93	3.31	-3.60	0.000
Students With Disabilities	-14.84	4.53	-3.27	0.001
Limited English Proficiency	-18.39	5.58	-3.29	0.001
Former Limited English Proficiency	2.91	20.94	-1.77	0.076
Third Grade NWEA-MAP Fall Reading Score	2.00	0.12	16.73	0.000
Third Grade NWEA-MAP Winter Reading Score	2.47	0.12	20.51	0.000
Second Grade Chronic Attendance	-7.76	3.26	-2.37	0.018

Policy/Research Question. Aligning with the state and district focus on early literacy, TPS sought to determine predictors of third-grade OCCT Reading scaled scores. Of particular focus for the analysis was the inclusion of so-called action-oriented levers. Action-oriented levers were predictors of third-grade reading proficiency that could potentially be addressed by changes in school and district policy (e.g. attendance, behavior, benchmark test performance), as opposed to external predictors, such as student demographics. The district was also

interested in characterizing more fully the population of our students who were at risk for retention in 2013–14: *What were their performance trajectories throughout the 2013–14 academic year? Could schools use the 2013–14 predictors to learn to self-monitor their own progress with interventions and special programs throughout 2014–15?* Therefore, the policy questions central to this capstone data analysis in Phase I were as follows: What are the second-grade and beginning-of-the-third-grade key indicators that predict third-grade OCCT Reading proficiency at the end of the third-grade academic year? and How are Tulsa Public Schools students faring across these early grades? Additional, ongoing data questions as part of Phase II include: How best do we communicate these predictors to our K–2 teachers and improve data literacy for third-grade reading outcomes?

Project Scope and Timeline. TPS only began universal reading benchmark screening in 2013–14. Thus, the study scope of Phase I included a retrospective analysis of the 2013–14 cohort. This analysis was presented to district leadership in summer and fall of 2014. Although our ability to characterize action-oriented levers for third-grade academic outcomes was still nascent at the start of the 2014–15 academic year, our need to appropriately identify students potentially at risk for retention in 2014–15, develop appropriate interventions, and monitor their progress throughout the academic year meant that we could not wait for multiple cohorts of data. The urgency surrounding third-grade reading in light of the Reading Sufficiency Act led the district to rollout data literacy and trainings using the lessons learned in Phase I throughout 2014–15. With the support of key members of the district leadership and as part of a broader rollout of a new student achievement dashboard, the team moved ahead with plans to provide custom data views on third grade reading success predictors for K–3 teachers as part of our new teacher dashboard. The team also produced regular forecasts for district academic trajectories and expected populations of third graders at risk for retention in 2014–15, at the district and school levels, to enable members of the district's executive and site leadership to self-monitor progress. The team also worked with a portfolio of 10 schools serving third grade (with the key support from each schools' Instructional Leadership Director) to conduct a pilot professional development three times a year. The focus of the trainings was effective use of key early on-track indicators to conduct mid-year performance evaluations at the school level.

Third-grade OCCT performance indicators were preliminary released in May 2015 for the 2014–15 academic year. While this data is considered preliminary until the final release of state-certified scores in late September 2015, the district saw declines of approximately 12% over 2013–14 in the number of students scoring unsatisfactory on the third-grade OCCT in 2014–15. We also saw substantively greater than district-average level decreases in the percent of students at risk for retention in 2014–15 in four of the ten pilot elementary schools.

Methods. To determine predictors of third-grade reading achievement in the 2013–14 cohort, multi-stage hierarchical multiple regressions were conducted with OCCT scaled score as the dependent variable. Demographic variables, such as race/ethnicity, Individualized Education Program (IEP), ELL, former ELL, mobility, and FRL, were entered progressively in Stage 1. Action-oriented student variables (discipline, attendance, coursework, benchmark NWEA Assessments) were entered in Stage 2. Fall (roughly September of the 2013–14 third-grade cohort's third-grade year) and winter (roughly January of the 2013–14 third- grade cohort's third grade year) NWEA benchmark scores were entered. Second-grade NWEA scores were not available for the study cohort, as the NWEA MAP was administered universally to 97% of third graders starting only in 2013–14. However, post-hoc analyses of the relationships between second-grade spring (May 2014) and third-grade fall (September 2014) NWEA MAP scores in the 2014–15 third-grade cohort suggest that third-grade fall NWEA MAP scores are generally equivalent or slightly lower (reflecting summer academic loss) than second-grade spring NWEA scores. Thus, results were unlikely to change had second-grade NWEA scores been available for the 2013–14 cohort. School-level clustering of variability was treated as a random-effect.

All variables (Stage 1 and Stage 2) were retained in the final model if they 1) explained a significant ($p < .05$) proportion of the variability in third-grade OCCT scores and 2) decreased the overall model fit values for the Akaike Information Criterion and the Bayesian Information Criterion. The Akaike Information Criterion and Bayesian Information Criterion are two tools for parsimonious model selection. Lower values indicate better model fit.

Results. Descriptive statistics for the third-grade cohort, based upon final retention status, are presented in Table 7. (Note, for these analyses, students who scored unsatisfactory

but were promoted on a probationary basis were not significantly different from students who scored unsatisfactory and were ultimately retained.)

Table 7. Demographic Profiles of Students At Risk and Not At Risk for RSA Retention in the 201314 Cohort (Tulsa)

2013–2014 Third-Grade Cohort	English Language Learners	Students w/ Disabilities	Economically Disadvantaged	Black or African American	Hispanic or Latino
OCCT Unsatisfactory (At Risk for Retention)	28.5%	46.2%	93.7%	30.7%	37.2%
OCCT Satisfactory or Above (Not At Risk for Retention)	4.5%	11.3%	76.6%	19.4%	28.5%

Race/Ethnicity, IEP and ELL status were significant predictors of final third-grade outcomes. Also, at-risk students were more likely to be socioeconomically disadvantaged, receive special programming (ELL/SPED) and be a member of a traditionally underrepresented minority group. While good-cause exemptions exist for the most severely intellectually disabled, students who take the modified Oklahoma Alternate Assessment Program exam (OAAP), and those newly receiving ELL services within the first two years, the majority of at-risk students did not receive automatic good-cause exemptions, even with special services. Interestingly, the results do not suggest that previous grade retention was a significant predictor of third-grade reading outcomes, nor was mobility between schools within the 2013–14 school year. There was a marginally significant trend toward students formerly classified as ELL (but qualifying out) having higher third-grade reading outcomes.

Discipline, as reflected by the number of total suspensions, did not significantly predict third-grade reading outcomes. Chronic absenteeism (>= 10% absent rate) was a significant predictor of third-grade OCCT scores.

Because suspensions involve missing school days, and thus instructional time, we replicated the regression analysis using a single composite measure of chronic absenteeism comprised of the total number of school days missed for either attendance or discipline. In this measure, a student suspended once for three days and missing 10 days of school would be counted as chronically absent. This composite measure was retained in the final model, suggesting that school days missed for any reason were detrimental to early on-track outcomes. Table 8 shows attendance patterns for third graders by 2013–14 RSA (Oklahoma's Reading Sufficiency Act) at-risk status. The number of unsatisfactory/failing grades was not a significant predictor of third-grade reading outcomes, though this may reflect floor effects. The proportion of U/F grades given in many of our district schools in our sample was <=5%.

Table 8. Attendance Profiles of Students at Risk and Not at Risk for RSA Retention in the 2013–14 Cohort (Tulsa)

2013–14 Third Grade Cohort	Attendance (ADA) %	Chronic (>=10%) Absence %
OCCT Unsatisfactory (At Risk for Retention	92.9%	21.5%
OCCT Satisfactory or Above (Not At Risk for Retention)	95.1%	13.7%

NWEA MAP Reading scores in both fall and winter were significant and reliable predictors of third-grade reading achievement. See Table 9 and Figures 6 and 7 for details. As Table 3 shows, the achievement gap between 2013 RSA at-risk students and 2013–14 students not at risk for RSA retention was marked. The median NWEA RIT score for at-risk students was equivalent to the reading level of a first-semester first-grade student. (A RIT score of 162.8 represents on-grade level performance for beginning of year first-grade students, per NWEA RIT score grade-level norms.) Students who were not at risk for RSA retention at the end of 2013–2014, however, had median RIT scores at the beginning of third grade that were at grade-level

norms for third grade. (A RIT score of 192.1 represents on-grade level performance for beginning of year third-grade students, per NWEA RIT score grade-level norms.)

Table 9. Fall 2013 NWEA Initial Third-Grade RIT Score by Final RSA Status (Tulsa)

2013–14 Third Grade Cohort	NWEA Fall 2013 Median	NWEA Fall 2013 25th Percentile	NWEA Fall 2013 75th Percentile
OCCT Unsatisfactory (At Risk for Retention)	168	157	178
OCCT Satisfactory or Above (Not At Risk for Retention)	191	182	199

Figures 6 and 7 show scatterplots of the relationships between 1) individual students' fall third-grade NWEA MAP scores/2013–14 OCCT scores and 2) school median fall NWEA scores for all third graders /school median third-grade 2013–14 OCCT. Note that a scaled score of 700+ represents a satisfactory (proficient) or higher score. NWEA Beginning of Year RIT scores and the degree of growth between fall and winter ("Beginning of Year" to "Middle of Year") were both reliable predictors of final OCCT score. These regression results, coupled with the knowledge that students typically are expected to grow by 10 RIT points in one year of academic schooling (per NWEA norms), suggests that students who were ultimately at risk for RSA would have needed to make over three years of growth, or 300% of their NWEA expected growth for 2013–14, to have moved out of the RSA risk category!

Figure 6. Relationship between Third-Grade individual Fall 2013 NWEA MAP scores and Third-Grade Spring 2014 individual OCCT Reading scaled scores (Tulsa)

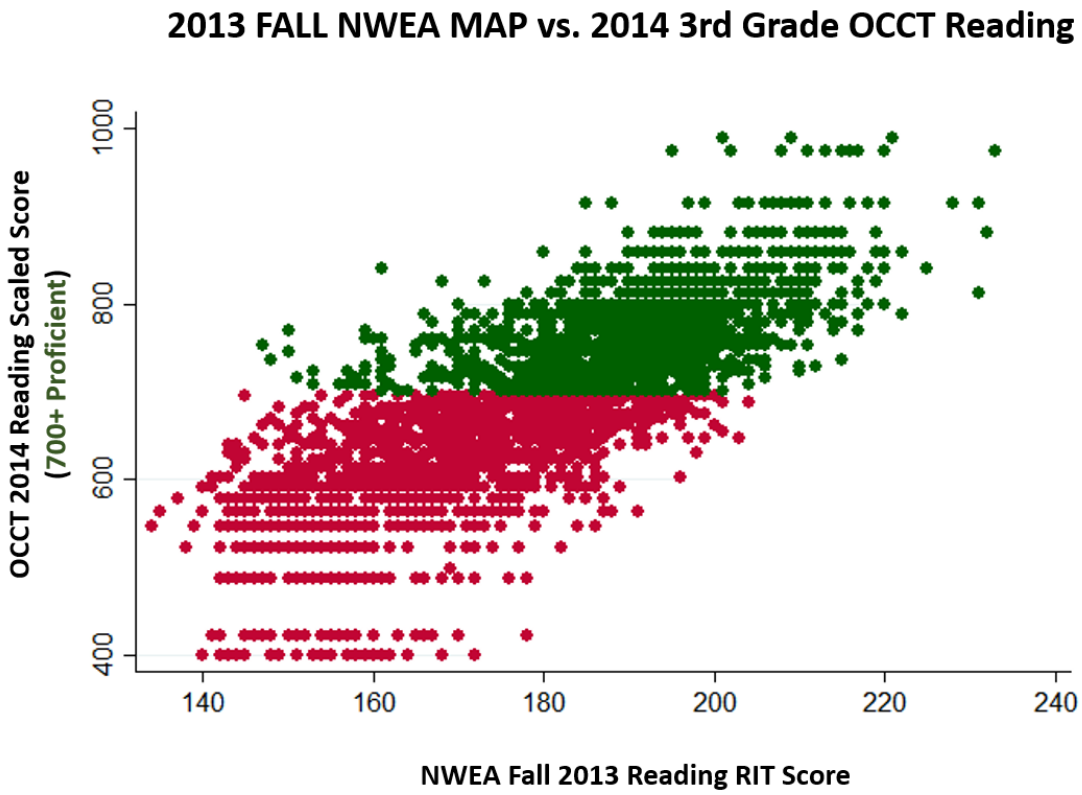
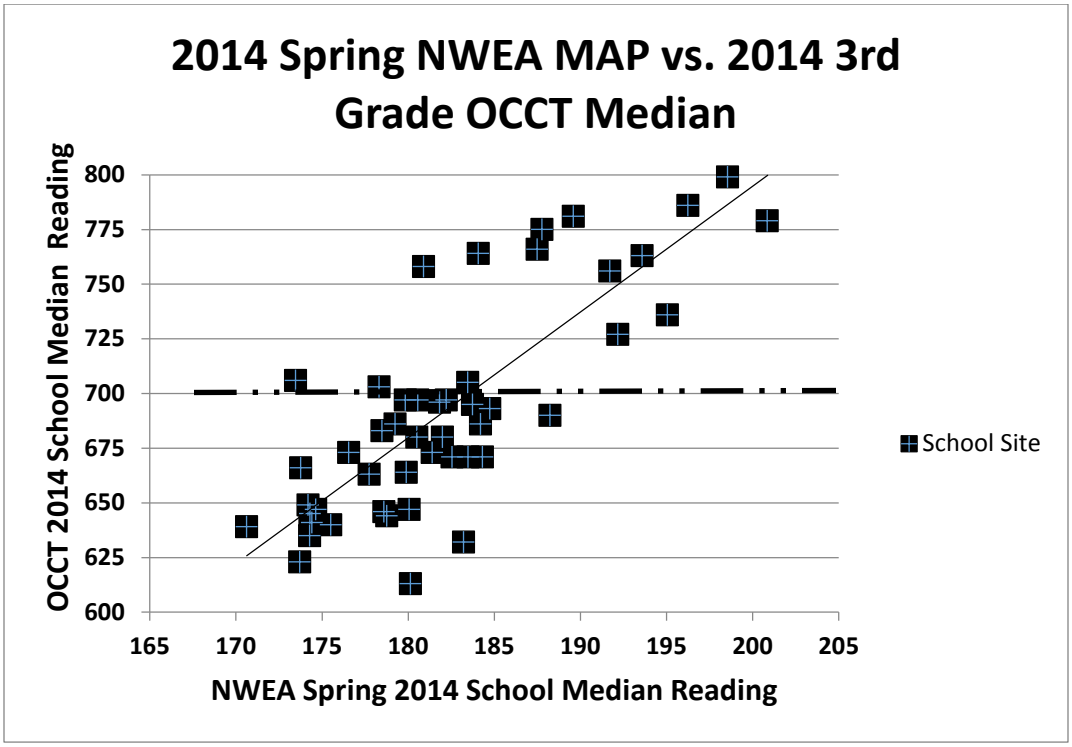


Figure 7. Relationship between Third-Grade school median Spring 2014 NWEA MAP scores and Third-Grade Spring 2014 school median OCCT Reading scores (Tulsa)



These startling growth requirements lead to an enhanced interest in utilizing the individualized expected growth goals provided by NWEA MAP to track the proportion of students making at least one year's academic growth. While the district average was 49%, at the school level, the analysis showed a range of 30%–83% of students meeting their yearly expected NWEA growth at various TPS schools.

Impacts and Key Takeaways . This capstone project summarizes one of the district's first opportunities to examine student achievement and growth data in early grades in a longitudinal and comprehensive manner. Third grade is a critical year with the introduction of the Oklahoma Reading Sufficiency Act. However, it is also the first year of state assessment. To address this potential gap, the Reading Sufficiency Act requires districts to track proficiency in early literacy throughout the K–2 period. Data from these analyses suggests that one year of tiered intervention in third grade is far too late—students who were ultimately at risk for RSA retention in 2013–14 looked academically more like first graders than third graders.

Additionally, 2013–14 expected growth goal attainment data suggests that the majority of second and third graders (>50%) were not even making a full year of academic progress in one year of academic instruction. This means that the majority of TPS at-risk students *not only* did not make the 300% of their expected growth on NWEA in one academic year they needed to truly move out of the risk category, but they majority fell further behind their not at-risk peers in the 2013–14 school year.

In summer 2014, the district also began parallel work on a new student achievement dashboard, termed Student STAT. The district leadership team determined that because NWEA MAP scores were the best predictors of third-grade reading success in our 2013–14 cohort, real-time data must be made available to teachers and principals showing 1) NWEA MAP scores for each child for each testing period in 2014–15 and 2) each student's percentage attainment of their own tailored growth goal, made available at the beginning of the year by NWEA MAP. In this manner, teachers would have the tools necessary to track each student's progress and to determine, via a combination of the child's most recent Reading RIT score and their attainment (or lack thereof) of reading growth, whether a child was likely to be at risk for retention in 2014–15. At the request of the deputy superintendent, the team conducted a technology and data survey prior to rollout of the new student dashboard with queries on the types of data our teachers need and prioritize for use in data-driven instruction. Only 23% of 556 district responders reported that they felt ready to leverage data effectively. Thus, while the team had empirical reason to believe it could identify students likely to be at risk for retention in the incoming 2014–15 school year, data knowledge, especially among K–3 teachers, was self-reported to be inadequate to use key at-risk data accurately.

To remedy this situation, the team began Phase II of the project. This phase continues to this time. Taking a three-pronged pilot approach toward improving data literacy and using the predictive power of NWEA Reading RIT scores, the team remained focused on improving K–3 instruction. First, the team designed and launched custom screens tracking benchmark NWEA MAP scores on the Student STAT dashboard. Appendix 1 shows an example of a training page from the Student STAT users' manual. This guide was distributed to all teachers in the district, and it included explicit instructions and suggestions for how to use NWEA cut scores and

expected growth numbers to determine if a student in K–3 was making adequate yearly progress, regardless of starting RIT score. Alongside initial training on the use of the Student STAT dashboard, conducted by the SDP Fellow and a team of school partners from the Department of Accountability (where the SDP Fellow is situated), all teachers at schools in TPS received data literacy training around the meaning of RIT scores and the use of expected growth during their introduction to Student STAT. Trainings were modeled after the ATLAS protocol for data from the School Reform Initiative. See the appendix for examples of a page from the Student STAT training manual which reviews NWEA growth goals, how to access them, and how to use them.

Second, the team produced formal forecasts of likely numbers of at-risk third graders by school after each benchmark assessment window. These forecasts were based upon empirically determined cuts cores from ROC curves, below which a child had a 75% or higher likelihood of scoring at risk on the 2014–15 OCCT. These reports were reviewed by the deputy superintendent, Chief Academic Officer, Chief Accountability Officer, and by Instructional Leadership Directors (ILDs) with their portfolios of 10–14 schools. One Instructional Leadership Director teamed with the SDP Fellow in an additional pilot project. In the ILD portfolio structure, principals of explicitly grouped schools meet regularly in portfolio-day professional development. As part of the 2014–15 pilots, worksheets were developed by the ILD for use with the pilot portfolio schools. These worksheets explicitly guided principals in determining exactly how many students needed to grow—and by how much—to "move the needle" on at-risk third-grade performance in a given school. The SDP Fellow attended portfolio meetings and provided regular data reports and technical assistance to principals in filling out the worksheets requested by the ILD. While results for 2014–15 OCCT scores are not publicly available until late September 2015, the team was encouraged in this pilot work by the fact that four of the 10 schools in this portfolio showed preliminary decreases in the number of at-risk (Unsatisfactory scoring) third-grade students in results provided to Tulsa Public Schools of >=40%. These four schools represent half of the total number of schools (eight) in the district achieving this magnitude of declines in numbers of at-risk students and well exceed the district's preliminary

12% reduction over 2013–14. See the appendix for examples of NWEA pilot ILD portfolio meeting materials.

Third, the team is explicitly using multiple outcome measures, both overall Reading RIT scores as well as the percentage of students meeting or exceeding their yearly growth goals, as outcome measures for program evaluation. The district's benchmark NWEA MAP Reading scores and the student's NWEA growth from fall to winter (the last NWEA test administered before the state exam) are reliable predictors of third-grade OCCT Reading performance. Therefore, NWEA benchmark scores are used to self-evaluate progress at the school and district level. Ongoing research and evaluation projects will explore the effectiveness of district interventions (and school-level intervention fidelity) at multiple time points throughout 2015–16. Professional development is under development by Curriculum & Instruction to guide teachers in the use of NWEA RIT scores to differentiate students into instructional small groups based upon benchmark test results after each administration throughout the early grades. Having the ability to forecast at multiple time points throughout the year which students are most at risk will enable us in future work to better match instructional techniques with a candidate pool of students who will be most benefited by targeted interventions and to help individual sites self-monitor progress.

Washington, DC

Context. Washington, DC, boasts the highest public Pre-K participation rates in the country with over 70% of three-year-olds and 95% of four-year-olds enrolled in programs within public schools and community based organizations. This allows for the unique opportunity to closely investigate the progression of Pre-K students into the early elementary years and to prepare these students for later points of proficiency measurement, specifically the third-grade measures of reading and mathematics. Given the unique alignment of Pre-K and K–12 data systems, DC can investigate the relationship between participating in Pre-K and later achievement outcomes.

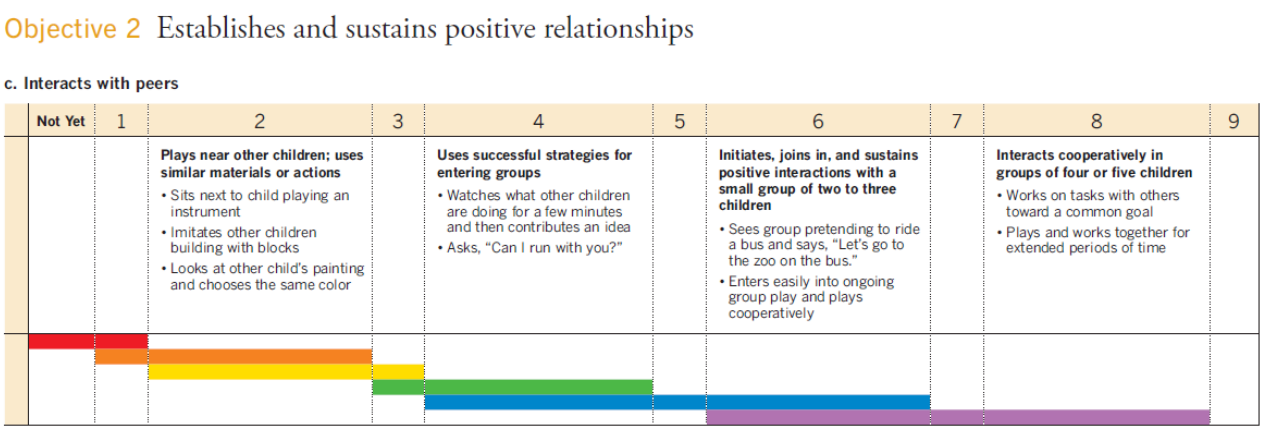
The Office of the State Superintendent of Education (OSSE) is the State Education Agency (SEA) for the District of Columbia and is charged with raising the quality of education for all DC residents. The District has a unique mixed delivery universal Pre-K program that serves

over 80% of all three- and four-year-olds in public schools and community-based organizations. The 2008–09 Pre-K Act brought together early childhood stakeholders with a common purpose of improving both access and quality of Pre-K and ensuring that students are Kindergarten ready. OSSE has additionally invested time and effort into establishing relationships with its sister agencies surrounding early childhood initiatives, including the Department of Human Services, Department of Health, and the Deputy Mayor's office. These entities also collaborate in overcoming challenges and obstacles.

Policy/Research Question. The District's near universal rates of Pre-K participation enable a deeper investigation of potentially predictive characteristics in the earlier grades than in other districts, while also allowing for an earlier examination of the impact on student outcomes as early as Kindergarten. Fewer than half of DC's third graders demonstrate proficiency in math or reading on the DC's Comprehensive Assessment System (DC-CAS). This project investigates factors that may support the development of targeted interventions to increase proficiency in third grade. Specifically, this project will examine the relationship between student-level characteristics in Pre-K and Kindergarten readiness to determine how to support learners and to ultimately increase third-grade reading and mathematical proficiency.

Project Scope and Timeline. The scope of this project examines the predictability of Pre-K student-level characteristics and the relationship to an identified Kindergarten assessment outcome variable: *Teaching Strategies GOLD.* The GOLD assessment is a developmentally appropriate teacher observation tool that measures students' progress along a continuum of 38 objectives aligned with the Common Core. The assessment also includes predictors of school success. The objectives fall into the following overarching domains: 1) Social Emotional, 2) Language, 3) Physical, 4) Cognitive, 5) Literacy, 6) Mathematics, 7) Science and Technology, 8) Social Studies, 9) The Arts), and 10) English Language Acquisition. Teachers record students' movement at three checkpoints (fall, winter, and spring) along this trajectory to ensure and support continued movement along the continuum. Figure 8 provides an example objective within the Social Emotional domain. A student would move from the Not Yet measure to an eight on the scale for full mastery of this objective when the student demonstrates interacting cooperatively in groups of four or five children.

Figure 8. Sample Objective Teaching Strategies GOLD (District of Columbia)



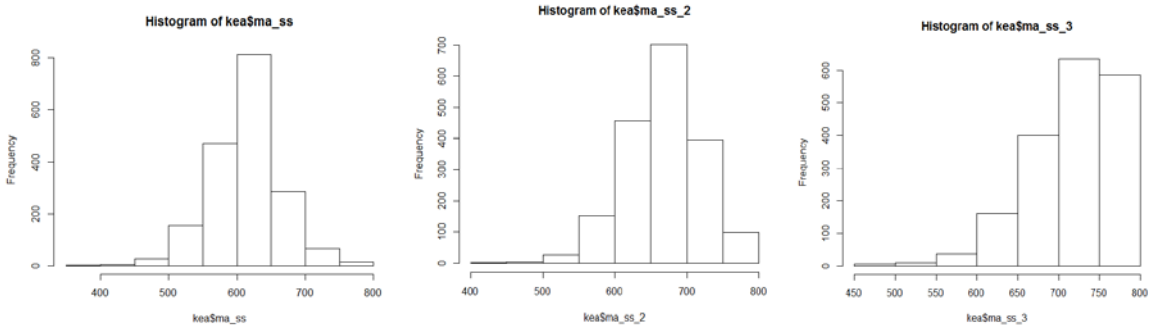
DC Public Schools (DCPS) first administered this tool in 2011–12 school year on a pilot basis, but it has since expanded to include all students in Pre-K4. This analysis explores Kindergarten readiness as measured in spring 2014 from the 2013–14 school year. Children included in this analysis will include the approximately 5,000 children enrolled in the Pre-K4 program at DCPS. From this sample, just over 2,000 students were also matched to their previous year of experience in school and were enrolled for two continuous years of Pre-K. Table 10 shows the demographic composition of this cohort and further divides the group into two cohorts: those students that participated in one year of Pre-K and those that participated in two full years of Pre-K. It is worth noting that the majority, or three out of four of the students, in the total cohort qualify for free and reduced lunch.

Table 10. Demographic Profiles of Students with One and Two Years of Pre-K Participation 2013–14 Cohort

2013–14 Pre-K Cohort	Male%	ELL%	SPED%	%FRL	% Black	% Hispanic
One Year of Pre-K Participation (*n*=1043)	52.73%	15.24%	4.21%	59.00%	50.05%	17.44%
Two Years of Pre-K Participation (*n*=2011)	50.22%	21.6%	14.12%	85.43%	68.87%	19.49%
TOTAL	51%	19.3%	10.77%	76.42%	62.44%	18.79%

The first exploration of data investigated differences in GOLD performance based exclusively on whether students had one or two years of Pre-K. For the purposes of this analysis, scaled scores were selected as the outcome variable ranging from a minimum of 200 and maximum of 800. The data received by DCPS included only raw and scaled scores, so this analysis should be treated as preliminary until a standardized data point is available, such as the NWEA score that is administered in both Tulsa and Rochester in the previous case studies. Figures 9(a) through (c) show the movement along the math domain in scaled scores at the three checkpoints, with distributions clustering around 600 in the fall and then the movement to 700 in the spring.

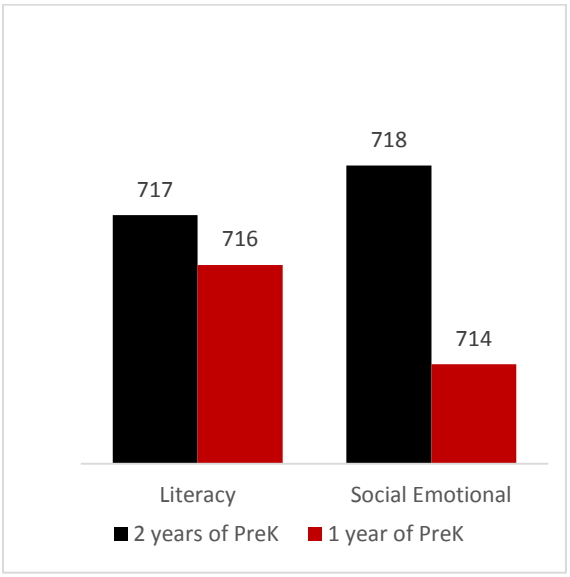
Figure 9. Distribution of Mathematical Kindergarten Readiness Scores during Fall, Winter, and Spring



9a. Fall 2013 **9b**. Winter 2013 **9c**. Spring 2014

An initial exploration using the GOLD scaled scores showed that children enrolled for two years outperformed their peers in both the literacy and social emotional domain on the GOLD assessment (Figure 10). However, these findings may be attributed to teacher familiarity with students and also rely on a comparison of scaled scores that are not standardized, as previously discussed.

Figure 10. One Year vs Two Years of PK—Impact on Kindergarten Readiness



To determine if any student characteristics predict GOLD assessment results, the following student-level data were included: Limited English Proficient Status, Students with Disabilities, Gender, Race, Ethnicity, Attendance and Discipline. The dataset was built by initially matching the dataset with the DCPS IDs to the district Unique Student Identifier. Following this, each additional dataset was merged to create a master dataset. For the preliminary model, the mathematics scaled scores were selected as the outcome variable as potential measurements of school readiness. Again, as this assessment is derived from teacher observation data and not a standardized measure, the math scaled score provided the best outcome measure at the time of the analysis, and results should be interpreted accordingly as exploratory until a standardized is available for analysis.

The majority of variables were binary (limited English, gender, and discipline), but categorical variables were also used for attendance and race. A linear regression model formula was used and run in R. When the math scaled score is selected over the composite score, the effect persists, although decreases for gender.

Table 11. Results of Regression (District of Columbia)

Indicators	B	*s.e.*	*T*	*P*
Intercept	782.375	3.953	197.932	<0.0000000000000002***
Limited English Proficient	-12.215	3.632	-3.363	0.00078***
Students with Disabilities	-45.755	3.065	-14.929	<0.0000000000000002***
Male	-5.268	1.89	-2.787	0.005352**
African American	-9.977	2.989	-3.338	0.000853***
Hispanic	-11.873	3.975	-2.987	0.00284**
Other	-1.013	4.991	-0.203	0.839226
(Reference Category: White)				
Frequently Absent	-6.982	4.95	-1.411	0.158424
Very Low Attendance	-11.811	8.932	-1.322	0.186153
Low Attendance	-2.064	9.563	-0.216	0.82914
Low Income	-12.517	2.182	-5.736	1.062E-08***
>1 Discipline Incident	-53.682	21.736	-2.47	0.013575*

Note. *N*= 3054. *F*(30.96, 3231) =42.04, p<.001. Adjusted R^2 = 0.1216.

Impacts and Key Takeaways. Table 11 shows all variables included in the model; the strongest predictors of school readiness were a student's limited English proficient status, race, and IEP status. While a relationship was found between attendance and the outcome variable, the relationship was weaker than with other variables. An additional exploratory regression including the Pre-K participation variable did not find a significant relationship with the outcome variable, likely due to a robust early intervention program that refers families of children with severe disabilities immediately to public Pre-K at the time of the child's third birthday. However, when children with IEPs were removed from the two Pre-K cohorts (one year of participation and two years of participation), a stronger relationship was found between students who participated in Pre-K for two years and the outcome variable. Similarly, a relationship between attendance and the outcome variable also emerged in this analysis. Both findings suggest a need for more comprehensive analysis and exploration.

The most pressing need for the District of Columbia is to identify an aligned, standardized assessment that can be used to explore these questions more deeply. Given the significant investment in early learning programming in the District and historic prioritization of early childhood policies over the past decade, this would enable policy makers to evaluate programs, inform future investments, and identify best practices, including action-oriented levers as discussed in the Rochester and Tulsa case studies. Attendance policies should continue to be developed that support higher rates of instructional time in Pre-K. Similarly, attendance policies prior to Pre-K should also be examined and potentially included in the child care Quality Ratings Improvement System to ensure that regular attendance patterns can be developed earlier. Developing a district wide on-track indicator will support teachers' and administrators' identification of those students who would benefit from targeted interventions.

Cross Agency Findings and Best Practices

All three agencies in this capstone project focused on finding indicators for early literacy. Specifically, the focus was on action-oriented predictors of third-grade literacy that were potentially modifiable at the school or district level, through modifications to academic policies and practices. While our contexts and outcome measures vary considerably, juxtaposing our findings presents a consistent and potentially encouraging story (Table 12). Together, we captured the age/grade span from three-year-olds in Pre-K to the end of third grade. Our most important finding is that there are key indicators, such as benchmark performance and growth, attendance, and quality Pre-K, which predict third-grade reading across districts in multiple geographic environments.

Table 12. Cross-Agency Findings Based on Predictive Modeling

Variables	Rochester, NY	Tulsa, OK	Washington, DC
End-of-Year Pre-K Assessment (Outcome)			√
Social Emotional Outcome (Outcome)			√
Third-Grade State Reading Test (Outcome)	√	√	
Discipline (Suspension)	√	√	
Attendance (Chronic Absenteeism)	√	√	√
Mobility		√	
Previous Program Participation			√
Gender	√		√
Race/Ethnicity	√	√	√
IEP (SWD)	√	√	√
ELL (LEP)	√	√	√
Former ELL (LEP)	√	√ - marginal	
Income (FRPL)	√	√	√
Coursework		√	
Homeless			√
Foster care			√
Subsidized Childcare Participant			√
Grade Retention	√	√	
District's Benchmark Tests (NWEA MAP Reading/Math)	√	√	
Kindergarten Child Observation Protocol	√		

Note: **√** denotes the variable used in the predictive analysis at a particular agency, and the variables highlighted in green show significance in the analysis.

Demographic variables, such as special program (IEP/ELL) status, are important predictors in several (though not all) of our case studies. We note that, although we did not term program participation directly as an action-oriented indicator, students in special education and English Language Learners face enormous challenges navigating the school

system in early grades and are consistently low-performing across agencies. While districts cannot change the needs of the students they serve, we caution that districts must be sensitive to the unique instructional needs of these populations. Potential policy implications of these findings include an increased need for tailored professional development for primary grade teachers who work with special populations, especially in high-poverty urban districts with high numbers of special learners. While these students often receive additional pull-out services with their SPED or ELL teachers, regular classroom teachers may struggle to adapt district- or school-specific reading interventions to these populations. So, too, might special services teachers struggle to mold the child's unique education support plans with their regular classroom instruction.

Although one case study (Tulsa) followed a group of students facing potential mandatory grade retention for poor performance in third grade, it is important to note that grade retentions as used in the models across the three agencies reflected decisions made at the district or school level. We did not directly test the impact of third-grade retention. Grade retention in K–2 did not seem to have a significant impact on reading proficiency outcomes. Qualitative questions arising from these findings focus on the consistency of policies and procedures for retentions. *How were retention decisions made in grades K–2, given that retained students were not significantly different from their peers in ultimate reading achievement?*

Coursework did not significantly predict third-grade reading performance. This represented a potentially significant deviation from high school indicators of on-track performance, wherein course grades, especially failed course grades in core subjects, are one of the strongest predictors of high school completion. This difference may be an artifact of the relatively low numbers of unsatisfactory/failing course grades issued by schools in the lower grades. Alternately, it may suggest broader policy implications surrounding consistency and expectations for grading within schools in the same district. Kindergarten to second grades are typically non-tested grades, and, lacking explicit guidance from state end-of-year exam standards and objectives, schools may struggle to provide uniformity in expectations. Further research would be desirable to determine the potential implications of this finding.

Attendance and/or chronic absenteeism were important predictors for early grade reading success. This was one of the most consistent findings of the cross-agency results. Two agencies did not find a unique effect of the number suspensions on early grade reading success. However, the cumulative number of days missed from school for any reason, including as a result of discipline, was significant. All agency results suggested that missing school, for any reason, severely hampered early readers. Thus, policy reviews around disciplinary decisions in the early grades should be sensitive to the lessons learned from chronic absenteeism. While overall attendance rates for a school are important (especially as multiple states may factor overall attendance into accountability report cards in some fashion), results from this capstone suggest that districts consider a tiered attendance intervention approach. Chronically absent students are a critical priority for K–3 attendance policies and planning. Interestingly, all agencies noted that attendance data in K–2, despite their predictive power, were most prone to data errors including missing data, incomplete data, or noted inconsistencies in applications across schools. This paradox is important to highlight. Anecdotally, principals across at least two agencies reported widely varying enforcement and oversight of attendance data. Attendance data matters, and it is worth districts' attention to develop policies and data architectures that support schools to track and act on patterns of absenteeism as they happen, before a students' attendance crosses the 90% threshold. The effects of missed school days lasts beyond a single school year.

Pre-K participation did show some positive impact on student outcomes. While the OSSE case study suggests that an additional year of Pre-K really matters, the RCS case study demonstrates the consistent and long-lasting effects of Pre-K on academic and behavioral outcomes in the three years following the intervention. Again, differences in results may reflect differences in the quality and content of Pre-K services across districts and across years of provided services.

Finally, cut scores for third-grade reading from benchmark tests are similar across agencies. Use of these empirically derived cut scores provides capabilities for detailed planning, and they may be a more effective way to communicate with practitioners than the results of complex predictive models. A data literacy survey in one agency suggested that, as districts

expand their overall modeling capabilities, a new knowledge gap may arise. The district central office may know with a high degree of precision which children are likely to be on track by third grade, but they may struggle to convey this complex knowledge in a digestible fashion to their K–3 teachers in the field. K–3 teachers may be particularly likely to be unfamiliar with concepts from benchmark testing, including expected growth or differences from grade-level norms. These teachers may benefit from a simpler-is-better approach. Lessons from initial pilot studies in both RCS and TPS suggest that districts must invest time and resources in explicitly training K–3 teachers in the use and value of data. Professional development, even for seemingly simple concepts, is crucial for ownership and widespread adoption of benchmark data. This is particularly true as both districts evaluating the use of NWEA benchmark scores experienced teacher and/or other stakeholder pushback around the use of K–2 adaptive testing during the studied time periods. These concerns were at least partially mitigated by hands on training in data literacy and regular communication of key data findings and the benefits of benchmarking in the early grades.

Perhaps the most promising use of predictive modeling in this context is the ability to flexibly adapt literacy stations, small-group read-alouds, and other forms of differentiated instruction to the demonstrated needs of the students. The goal of predictive modeling is, ultimately, to render its own predictions inaccurate, and the best possible use of early on-track indicators is to guide interventions such that, by the time a child reaches third grade, their academic and other needs have been appropriately met to ensure reading success. Teachers need tailored professional development on the use of data, and they need easy access to their own students' data, when they are making the intervention decisions that matter—typically far sooner than traditional reporting. One teacher recruited as part of the ongoing pilot projects of our agencies expressed it best: "We're [teachers are] the ones expected to do the most to change our students' on-track outcomes, but we have the least access to data to do it with." We believe these words speak for themselves regarding how best districts should proceed in using the outcomes of predictive modeling to support early grade reading success.

This capstone report supports an integrated approach based on early grade indicators, the building of a teacher-supportive environment, and policies at the school level to move the needle for early elementary performance. Possible next steps involve building early warning systems in reading using multivariate assessments of reading skills (Koon & Petscher, 2015), further investigating the potential positive effects of Pre-K intervention and reexamining grade retention policies and practices.

Appendices

Table 1. Results of OLS Regression Analysis after Lasso Method (Rochester)

Indicators	B	*s.e.*	*t*	*p*
Intercept	-63.30	13.63	-4.64	0.000
African American	-1.97	2.25	-0.87	0.382
Hispanic	-0.37	2.53	-0.15	0.883
Asian	8.44	5.14	1.64	0.101
Other	-9.55	11.47	-0.83	0.405
(Reference Category: White)				
Male	-2.27	1.23	-1.85	0.064
Poverty	-2.52	2.43	-1.03	0.301
Students With Disabilities	-5.34	1.97	-2.71	0.007
Limited English Proficiency	-5.35	2.65	2.02	0.043
Former Limited English Proficiency	-8.73	4.58	1.91	0.057
Ever Retained by the End of 3rd Grade	-5.93	4.76	-1.25	0.213
Pre-K Participation	-0.68	1.34	-0.51	0.609
Kindergarten End-of-Year Progress Assessment	1.29	0.94	1.37	0.172
1st-Grade NWEA-MAP Spring Reading Score	0.35	0.08	4.48	0.000
2nd-Grade NWEA-MAP Spring Reading Score	1.04	0.08	12.59	0.000
1st-Grade NWEA-MAP Spring Math Score	0.26	0.09	3.09	0.002
2nd-Grade NWEA-MAP Spring Math Score	0.19	0.08	2.44	0.015
2nd-Grade Attendance	22.31	12.02	1.86	0.064
Kindergarten Attendance	-12.53	9.44	-1.33	0.185
Kindergarten Suspension	6.24	3.30	1.89	0.059
School Fixed Effects(Omitted)				

Note. N= 1451. $F(48, 1402) = 42.04$, $p<.001$. Adjusted $R^2 = .576$.

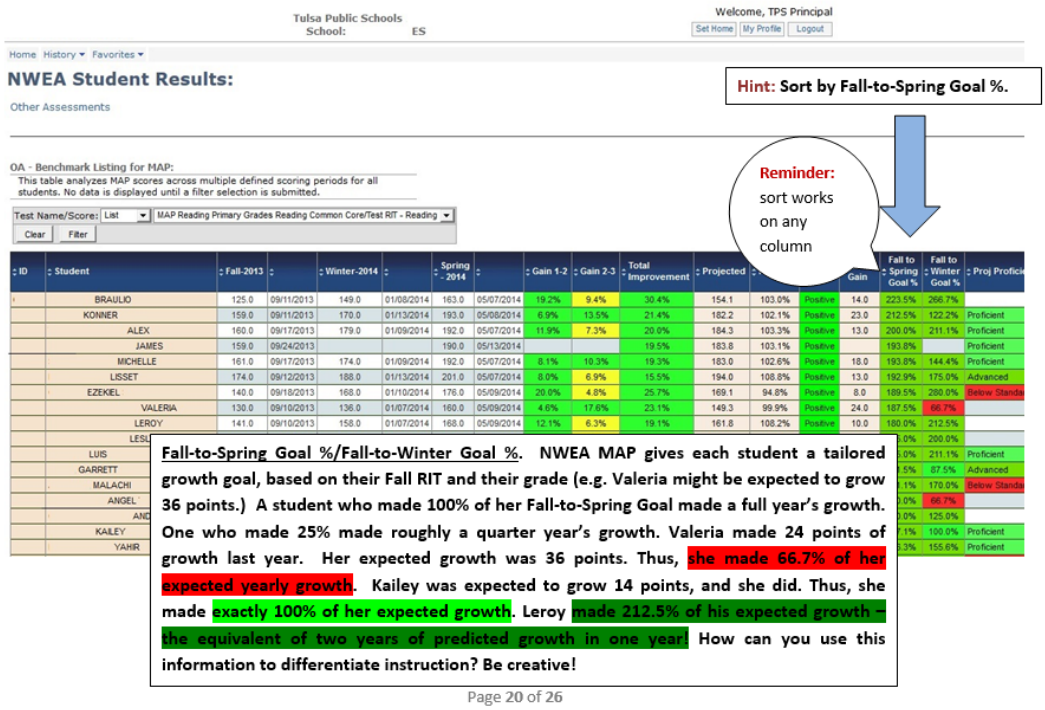
Table 2. ROC Curve Analysis Sensitivity and 1-Specificity Results (Rochester)

	Composite Success Factor		**2nd-Grade Reading**		**2nd-Grade Math**		**1st-Grade Reading**		**1st-Grade Math**		**2nd-Grade Attendance (Marginal Sig.)**	
Percentiles for Predictor	1-Specificity	Sensitivity	1-Specificity	Sensitivity	1-Specificity	Sensitivity	1-Specificity	Sensitivity	1-Specificity	Sensitivity	1-Specificity	Sensitivity
10th	89.23%	100.00%	88.53%	99.10%	88.81%	100.00%	88.75%	99.03%	89.04%	99.03%	88.96%	97.35%
20th	78.54%	100.00%	76.75%	99.10%	78.77%	100.00%	76.99%	99.03%	76.87%	99.03%	78.90%	94.69%
30th	67.85%	100.00%	67.28%	99.10%	67.95%	100.00%	67.01%	99.03%	67.39%	98.06%	66.67%	86.73%
40th	57.15%	100.00%	55.50%	98.20%	56.59%	99.10%	57.50%	99.03%	55.86%	98.06%	57.12%	84.07%
50th	46.53%	98.95%	47.24%	97.30%	46.72%	96.40%	45.21%	97.09%	45.92%	98.06%	47.06%	76.99%
60th	35.99%	96.84%	35.77%	95.50%	36.10%	94.59%	36.33%	93.20%	36.68%	96.12%	38.60%	63.72%
70th	25.37%	95.79%	24.30%	94.59%	25.96%	88.29%	25.03%	89.32%	23.02%	91.26%	28.38%	51.33%
75th	---	---	19.73%	92.79%	19.39%	85.59%	19.67%	83.50%	20.90%	85.44%	23.07%	39.82%
80th	14.82%	93.68%	15.36%	89.19%	15.29%	80.18%	15.74%	80.58%	15.10%	76.70%	17.13%	30.09%
85th	9.88%	89.47%	---	---	---	---	---	---	---	---	---	---

90th	5.46%	74.74%	6.00%	66.67%	6.25%	55.86%	6.46%	59.22%	6.08%	58.25%	6.60%	10.62%
Tentative cutoff	289		185		189		171		172			

Appendix 2: Student STAT training manual page for NWEA scores and expected growth (Tulsa)

Data Page 4, Part 4: Select the NWEA Student Results link from the left navigation menu. You can sort on any column by clicking on the header!



Appendix 3: Example worksheet for pilot NWEA data literacy schools (Tulsa)

NWEA
Northwest Evaluation Association

MAP Reading: Percent Meeting Projected Growth

School	Kindergarten						First Grade						Second Grade						Third Grade					
	2013-2014	Winter Goal %	Winter Number	Winter Actual Goal %	Spring Goal %	Spring Number	2013-2014	Winter Goal %	Winter Number	Winter Actual Goal %	Spring Goal %	Spring Number	2013-2014	Winter Goal %	Winter Number	Winter Actual Goal%	Spring Goal %	Spring Number	2013-2014	Winter Goal %	Winter Number	Winter Actual Goal%	Spring Goal %	Spring Number
1	29.4						44.9						56.6						64.1					
2	34.9						52.6						39.5						33.3					
3	63.6						49.5						29.1						56.7					
4	20.2																							
5	61.4																							
6	63.8						83.3						31.5						78.8					
7	50.0						78.8						78.0						70.4					
8	53.6						57.1						49.4						59.2					
9	23.9						36.5						26.8						50.0					
10	43.3						59.2						39.7						38.7					
11	33.3						43.2						72.0						47.5					

Guidance for Setting Goals That Will Result in at Least 50% of Your Students Meeting Projected Growth Goals

Range of 0-29= RED	Range of 30-44= ORANGE	Range of 45-64= YELLOW	Range of 65-100= GREEN
Aim to project at least 30 more % points	Aim to project at least 20 more % points	Aim to project at least 10 more % points	Aim to project at least 5 more % points

★ We are going to engage in activities that will allow you to determine a projected SPRING Score for your students. We will also determine the number of students needed to meet your target. ★ Review your numbers above and make reasonable projections for each grade. By meeting projected growth goals, you add value to students.

Projection	Kindergarten	First Grade	Second Grade	Third Grade	Fourth Grade
Winter					
Spring					

★ Once you determine your projection, you will need to strategically plan how you are going to accomplish your goal.

- ✓ *What new learning will you need to acquire?*
- ✓ *What new learning will your teachers need to acquire?*
- ✓ *How will you support teachers with planning for Guided Reading and Literacy Centers based on RIT scores?*
- ✓ *How will you ensure teachers are aiming to differentiate lessons/activities to meet the needs of their students grouped by RIT ranges?*

References

Aguiar, Everaldo, et al. "Who, when, and why: a machine learning approach to prioritizing students at risk of not graduating high school on time." *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*. ACM, 2015.

Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education*, *70*(2), 87–107.

American Institutes for Research, (2013) *Technical Descriptions of Risk Model Development:* Early and Late Elementary Age Groupings (Grades 1–6).

Barfield, K. A., Hartman, J., & Knight, D. (2012). Early Warning Systems: It's Never Too Early. T H E Journal, 39(2), 18–19.

Barry, M., & Reschly, A. L. (2012). Longitudinal Predictors of High School Completion. School Psychology Quarterly, 27(2), 74–84. doi:10.1037/a0029189

Bowers, A.J., Sprott, R.*, Taff, S.A.* (2013) Do we Know Who Will Drop Out? A Review of the Predictors of Dropping out of High School: Precision, Sensitivity and Specificity. The High School Journal, 96(2), 77–100. doi:10.1353/hsj.2013.0000

Campbell F.A., & Ramey C.T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development, 65,* 684–698.

Davis, M., Herzog, L., & Legters, N. (2013). Organizing Schools to Address Early Warning Indicators (EWIs): Common Practices and Challenges. Journal Of Education For Students Placed At Risk, 18(1), 84–100. doi:10.1080/10824669.2013.745210

Desimone, L. M., Long, D. (2010). Teacher Effects and the Achievement Gap: Do Teacher and Teaching Quality Influence the Achievement Gap Between Black and White and High- and Low-SES Students in the Early Grades? Teachers College Record, 112(12), 3024–3075.

Ehrlich, B., Gwynne, J., Pareja, A.S., and Allensworth, E., (2013). Preschool Attendance in Chicago Public Schools UChicago CCSR. The University of Chicago Consortium on Chicago School Research.

Frazelle, A (2015). Practicioner's Guide to Implementing Early Warning Systems. IES.

Hiebert, E.H., Pearson, D. P. (2010). National Reports in Literacy: Building a Scientific Base for Practice and Policy. Educational Researcher, 39(4), 286–294.

Hu, Y., Lo, C., & Shih, S. (2014). Developing early warning systems to predict students' online learning performance. Computers in Human Behavior, 469–478. doi:10.1016/j.chb.2014.04.002

Henry, K., Knight, K., & Thornberry, T. (2012). School Disengagement as a Predictor of Dropout, Delinquency, and Problem Substance Use During Adolescence Early Adulthood. Journal Of Youth & Adolescence, 41(2), 156–166. doi:10.1007/s10964-011-9665-3

Knowles, J. E., (2014). Of Needles and Haystacks: Building an Accurate Statewide Dropout Early Warning System in Wisconsin, Wisconsin Department of Public Instruction.

Koon, S., & Petscher, Y. (2015). Comparing methodologies for developing an early warning system: Classification and regression tree model versus logistic regression, Applied Research Methods report for Institute of Education Sciences (IES), accessed on April 22, 2015 from: <http://ies.ed.gov/pubsearch/pubsinfo.asp?pubid=REL2015077>

Leonard N. Masse and W. Steven Barnett. (2002). A Benefit-Cost Analysis of the Abecedarian Early Childhood Intervention, New Brunswick, N.J.: National Institute for Early Education Research. <http://nieer.org/resources/research/AbecedarianStudy.pdf>

Mac Iver, M., & Messel, M. (2013). The ABCs of Keeping on Track to Graduation: Research Findings from Baltimore. Journal Of Education For Students Placed At Risk, 18(1), 50–67. doi:10.1080/10824669.2013.745207

Neild, R., Balfanz, R., & Herzog, L. (2007). An Early Warning System. Educational Leadership, 65(2), 28–33.

Neild, R. (2009) Falling Off Track during the Transition to High School: What We Know and What Can Be Done.

Sapers, J. (2014). Dropping Out: Is Your First Grader at Risk? Accessed on June 29, 2015 <https://www.gse.harvard.edu/news/ed/14/01/dropping-out-your-first-grader-risk>

Shonkoff, J.; Phillips, D. A.; Council, N. R. (Eds.). (2000). From Neurons to Neighborhoods: The Science of Early Childhood Development. Washington, DC: National Academy Press.

Soland, J. (2013). Predicting High School Graduation and College Enrollment: Comparing Early Warning Indicator Data and Teacher Intuition. Journal of Education for Students Placed At Risk, 18(3/4), 233–262. doi:10.1080/10824669.2013.833047

Tang (2014). Adolescent Pregnancy's Intergenerational Effects: Does an Adolescent Mother's Education Have Consequences for Her Children's Achievement? Journal of Research on Adolescence.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of Royal Statistical Society Series B., Vol. 58, No. 1, pages 267–288.

Use Early Warning Systems to Keep Students in School. (2011). What's Working, 3(6), 6.

West, Thomas C. (2013) Just the right mix: Identifying Potential Dropouts in Montgomery County Public Schools Using an Early Warning Indicators approach. *Montgomery County Public Schools*.