The Magnitude and Correlates of Teacher Effectiveness in Early Grades Reading

In this research brief, we leverage statewide student achievement data to better understand the magnitude and correlates of teacher effectiveness in early grades reading. Our analyses return several key takeaways.

(1) Teacher effects in early grades reading are large. This motivates continued efforts to ensure that teachers have the knowledge, skills, and resources necessary to succeed.

(2) There is a need to better understand effective literacy preparation and to possess more granular measures of high-quality preparation practices.

(3) Proxies for teacher knowledge—Foundations of Reading scores, possessing a reading license—predict teacher effectiveness on the TRC but not on DIBELS. It is important to understand these across-assessment differences to inform policy decisions.

Introduction

Early literacy skills are essential to later academic success, as students who are not proficient readers by the end of third grade go on to have higher rates of behavioral incidents, grade retention, and school dropout. Given the importance of early literacy, many states have taken concerted steps to improve K–3 students' reading ability. North Carolina is no exception. In 2012, the state General Assembly enacted *Read to Achieve*, a K–3 reading policy that (1) required third grade students to be proficient readers and (2) supported this requirement through formative and on-going reading assessments to help teachers and schools target instruction.

These reading assessments measure foundational reading skills, fluency, and comprehension and provide consistent, statewide data on reading achievement in early grades. These data can meaningfully contribute to decisions of policy and practice at the school, district, and state levels.

As the largest single preparer of teachers for the state's public schools, the University of North Carolina (UNC) System is also focused on the importance of early literacy and the role of teachers in developing proficient readers. Towards this end, the UNC System commissioned joint research by the Education Policy Initiative at Carolina (EPIC) at UNC Chapel Hill and the Cato College of Education at UNC Charlotte. Broadly, this commissioned research focuses

on teacher effectiveness in early grades reading and the instructional resources and practices used by early grades reading teachers. In this research brief, we report findings from our teacher effectiveness analyses. In particular, we answer the following questions for early grades reading teachers: (1) How large are teacher effects in early grades reading? (2) Do measures of teacher preparation predict teacher effectiveness? and (3) Do teacher exam scores and certification areas predict teacher effectiveness? These analyses identify areas for further research, reinforce and challenge current policy, and motivate continued action to ensure that teachers have the knowledge, skills, and resources to support students' literacy development.

Background

We assess teacher effectiveness in early grades reading by analyzing four years of student test scores (2014-15 through 2017-18) from the mCLASS Reading 3D exam.[1] It is important to note that test scores are only one measure of teacher effectiveness and that a variety of factors (e.g., reading behaviors and resources at home, curriculum and resources provided to teachers) influence student achievement in reading. mCLASS is a literacy assessment administered to students in grades K–3 at the beginning, middle, and end of the school year. The exam measures developing reading skills with two main assessments: DIBELS (Dynamic Indicators of Basic Early Literacy Skills) and TRC (Text Reading and Comprehension). DIBELS includes up to six standardized sub-tests to measure phonics, phonemic awareness, and fluency. Students receive a DIBELS composite score that corresponds to a criterion-referenced reading level of above, below, or well below benchmark. The TRC assesses concepts of print, reading behaviors, word recognition, and the ability to read full text passages with fluency and accuracy and answer comprehension questions. Teachers assign students to a TRC reading level (e.g., RB, PC, A–U), which serves as a grade-level benchmark of proficiency.

From these mCLASS data, we created two outcome measures. For DIBELS, our outcome measure is students' end-of-year composite score, standardized by grade and year. To better isolate teachers' contributions to the end-of-year composite score, we control for students' beginning-of-year composite score in analyses (standardized by grade and year). For the TRC, our outcome measure is the number of reading levels that a student improved from the beginning to the end of the school year. For example, students starting first grade at level 'D' and ending first grade at level 'J' improved six levels. Because the change in reading levels may be influenced by the level at which students start,[2] we compare the growth of students to that of other students starting the year at the same reading level.

To address our first research question, the size of teacher effects in early grades reading, our sample includes all reading teachers-of-record in grades K–3. We estimate individual teacher effectiveness with a multi-level model— students nested within teachers and schools—that controls for a rich set of student, classroom/peer, and school covariates.[3] For our second and third research questions, considering the associations between teacher effectiveness and teacher preparation, exam scores, and certification areas, we limit our sample to first, second, and third-year teachers. We focus on these teachers because the effects of preparation and licensure are most relevant in the early-career period. For these preparation and licensure analyses, we prefer multi-level models that make teacher effectiveness comparisons across all schools. We present results from these multi-level models in Figures 3-7. We supplement these analyses with fixed effect models that limit comparisons to teachers in the same school. We reference, but do not display, results from these fixed effect models.[4] In all of the analyses for our second and third research questions, we control for a robust set of student, classroom/peer, teacher, and school covariates.
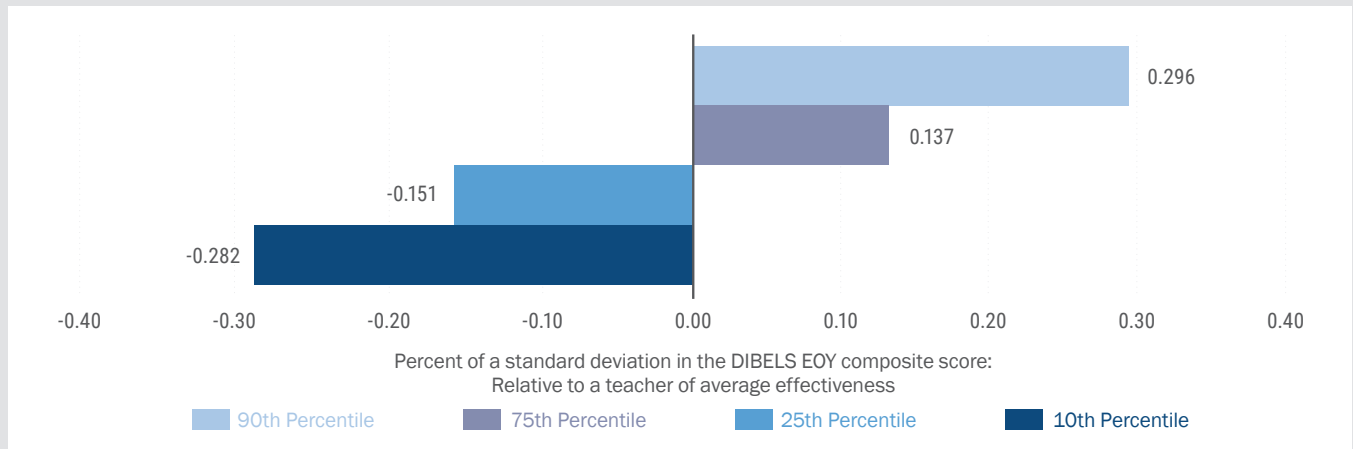
---

[1] We use mCLASS data because it was the official, statewide assessment for K–3 literacy from the 2013-14 through 2018-19 academic years.

[2] Students' progression through lower reading levels may be quicker than through upper reading levels given the skills required to progress and given grade-level expectations for growth.

[3] Our measure of individual teacher-year effectiveness is the teacher-level random effect generated by this model.
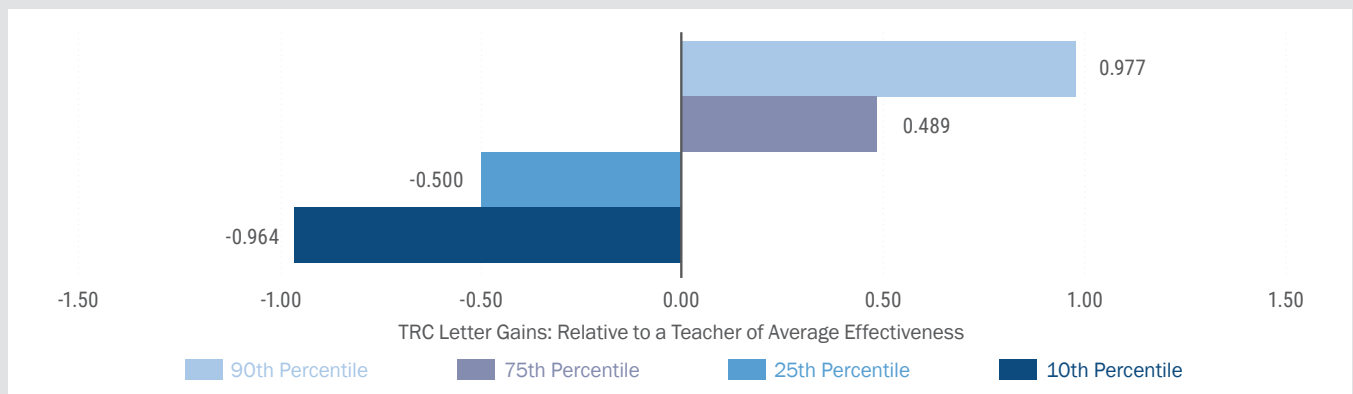
[4] Please see https://publicpolicy.unc.edu/files/2020/06/Literacy_Results_Packet_Final_6-22-20.pdf for the full set of our DIBELS and TRC results.

Figure 1: The Impact of Teachers on DIBELS Composite Scores

Percent of a standard deviation in the DIBELS EOY composite score:
Relative to a teacher of average effectiveness

■ 90th Percentile   ■ 75th Percentile   ■ 25th Percentile   ■ 10th Percentile

Note: *This figure displays the effectiveness of individual teachers at the 10th, 25th, 75th, and 90th percentile of DIBELS effectiveness relative to the average teacher. Individual teacher effectiveness estimates come from models in which K-3 students' end-of-year DIBELS composite score (standardized within grade and year) is the outcome. The sample includes all teachers in the 2014-15 through 2017-18 school years who are linked to DIBELS data. Models control for student, classroom, and school covariates.*

Figure 2: The Impact of Teachers on TRC Reading Levels

TRC Letter Gains: Relative to a Teacher of Average Effectiveness

■ 90th Percentile   ■ 75th Percentile   ■ 25th Percentile   ■ 10th Percentile

Note: *This figure displays the effectiveness of individual teachers at the 10th, 25th, 75th, and 90th percentile of TRC effectiveness relative to the average teacher. Individual teacher effectiveness estimates come from models in which K-3 students' reading level gain on the TRC is the outcome. The sample includes all teachers in the 2014-15 through 2017-18 school years who are linked to TRC data. Models control for student, classroom, and school covariates.*

How large are teacher effects in early grades reading?

Previous analyses show that teachers have large effects on student achievement. In fact, among school-related factors, teachers matter the most to student learning. As much of the prior work on teacher value-added has focused on teachers in grades 4-12, we performed comparable analyses in early grades reading. These analyses provide one way to quantify the importance of teachers and can motivate policy efforts to ensure that teachers have the knowledge, skills, and resources to support students' literacy development.

Figure 1 displays the effectiveness of individual teachers at the 10th, 25th, 75th, and 90th percentiles of DIBELS effectiveness. Relative to students taught by an average teacher (at the 50th percentile of DIBELS effectiveness), Figure 1 shows that students taught by a highly-effective teacher (at the 90th percentile) have DIBELS composite scores nearly 30 percent of a standard deviation higher. This finding is equivalent to end-of-year composite scores 26 points higher in first grade and 32 points higher in second grade. The difference between a highly effective and an ineffective teacher is twice as large. Students taught by a teacher at the 90th percentile of DIBELS effectiveness have DIBELS composite scores nearly 60 percent of a

standard deviation higher than comparable students taught by a teacher at the 10th percentile.

Figure 2 displays the effectiveness of individual teachers at the 10th, 25th, 75th, and 90th percentiles of TRC effectiveness. Relative to students taught by an average teacher (at the 50th percentile of TRC effectiveness), Figure 2 shows that students taught by a highly–effective teacher grow one more TRC reading level between the beginning and end of the school year. To put this result into perspective, we note that, on average, students gain five TRC reading levels in first grade and four reading levels in second grade. Likewise, students taught by a teacher at the 90th percentile of TRC effectiveness grow nearly two more reading levels between the beginning and end of the school year than comparable students taught by a teacher at the 10th percentile.
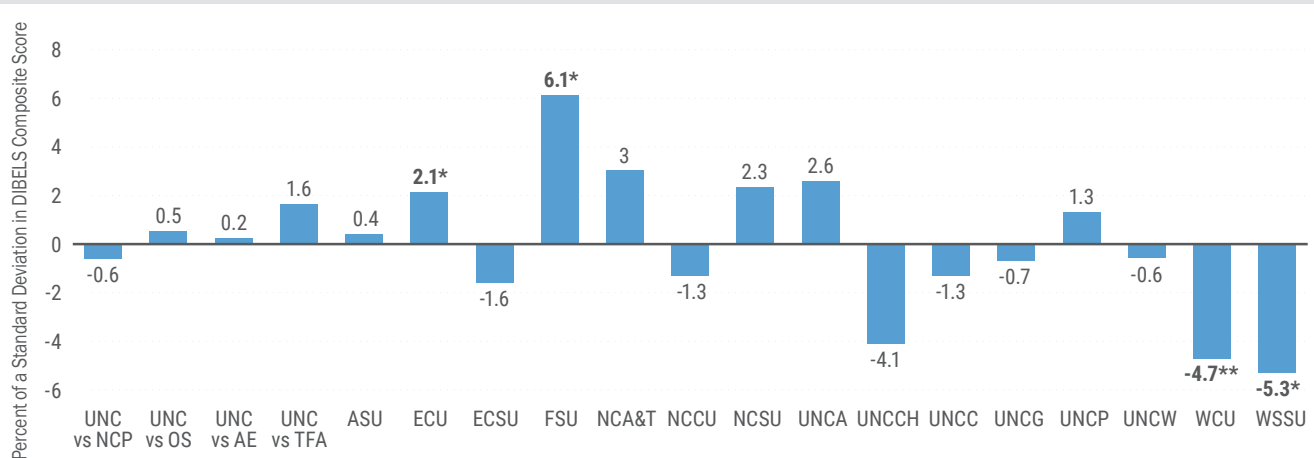
Do measures of teacher preparation predict teacher effectiveness?

In these analyses, we focus on three measures of teacher preparation: (1) a teacher's route of preparation; (2) the UNC System institution that a teacher attended (if applicable); and (3) ratings of the early literacy instruction at teacher preparation programs.[5] The first and second measures indicate whether teachers prepared through certain routes or at certain UNC System institutions are more effective. The third measure indicates whether graduates of preparation programs with more highly–rated literacy instruction are more effective than graduates of preparation programs with lower ratings.

Figure 3 presents DIBELS results for teacher preparation routes and programs. Regarding preparation routes, we find that early-career teachers from the UNC System are no more or less effective than in–state private, out-of-state, alternative entry, or Teach For America (TFA) teachers. Two UNC System institutions have positive DIBELS results. Students taught by early-career teachers prepared at ECU and FSU have adjusted-average DIBELS scores 2.1 and 6.1 percent of a standard deviation higher, respectively, than students taught by non-UNC System prepared teachers. Notably, these two programs also had positive results in EPIC's most recent analyses of fourth and fifth grade reading scores (EOG exams).[6] Conversely, two UNC System institutions, WCU and WSSU, have negative DIBELS results in comparison to non–UNC System prepared teachers. None of the DIBELS estimates for ECU, FSU, WCU, and WSSU remain statistically significant in models comparing teacher effectiveness within schools.

Figure 3: DIBELS Results for Teacher Preparation Routes and Programs (Statewide Teacher Effectiveness Comparisons)
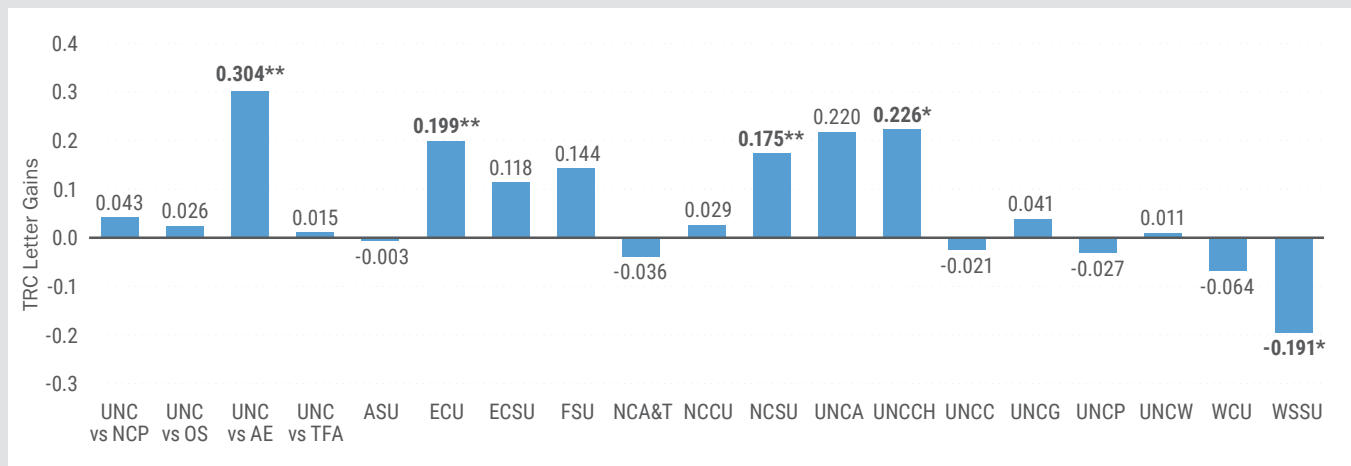

Note: *This figure displays results from models in which K-3 students' end-of-year DIBELS composite score (standardized within grade and year) is the outcome. The sample includes early-career teachers (<3 years of experience) in the 2014-15 through 2017-18 school years. Models control for student, classroom, teacher, and school covariates. '*' and '**' indicate statistical significance at the 0.05 and 0.01 levels, respectively. NCP=NC private institutions; OS=out-of-state; AE=alternative entry; TFA=Teach For America.*

[5] These ratings come from the National Council on Teacher Quality (NCTQ). NCTQ assesses whether elementary teacher preparation programs meet standards for literacy instruction in phonemic awareness, phonics, fluency, vocabulary, and comprehension. Based on these ratings, NCTQ scores each preparation program from 0–4 on their Early Literacy standard. See https://www.nctq.org/dmsView/2018_Reading_Findings for more on these ratings.

[6] These analyses used data from 2012–13 through 2016–17 and focused on teachers with less than five years of experience.

Figure 4: TRC Results for Teacher Preparation Routes and Programs (Statewide Teacher Effectiveness Comparisons)
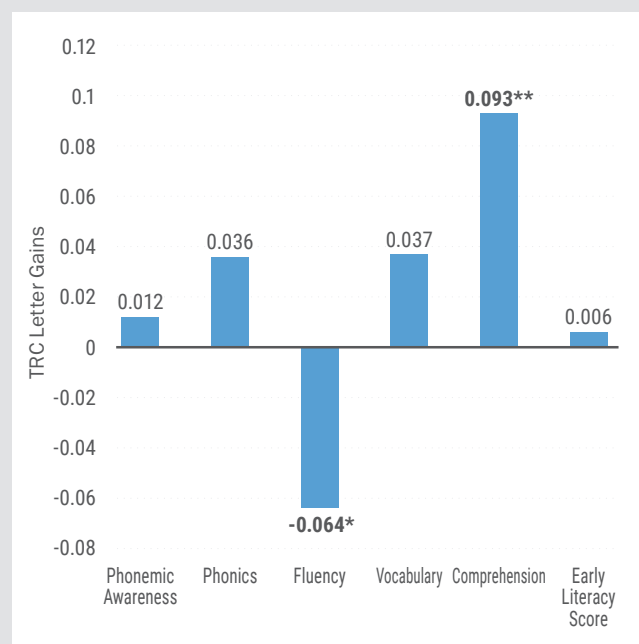
Note: *This figure displays results from models in which K-3 students' letter gain on the TRC is the outcome. The sample includes early-career teachers (<3 years of experience) in the 2014-15 through 2017-18 school years. Models control for student, classroom, teacher, and school covariates. '*' and '**' indicate statistical significance at the 0.05 and 0.01 levels, respectively. NCP=NC private institutions; OS=out-of-state; AE=alternative entry; TFA=Teach For America.*

Figure 4 presents TRC results for teacher preparation routes and programs. Regarding preparation routes, we find that early-career teachers from the UNC System are more effective than their alternative entry peers. In particular, students taught by an early-career UNC System prepared teacher grow nearly one-third of a TRC reading level more than comparable students taught by an alternative entry teacher. Three UNC System institutions have positive TRC results. Relative to non-UNC System prepared teachers, students taught by graduates of ECU, NCSU, and UNCCH make larger TRC reading gains. Conversely, students taught by WSSU graduates make smaller TRC reading gains. ECU and WSSU are the only two UNC System institutions with statistically significant results in both the DIBELS and TRC analyses.[7]

To extend our preparation analyses, we assess whether NCTQ's ratings of the literacy instruction in elementary teacher education programs are associated with the effectiveness of graduates from in-state institutions (public or private). Figure 5 displays results from our TRC analyses; since only one DIBELS result is statistically significant, we comment on, but do not display those findings. Regarding the five indicators of scientifically-based reading instruction, Figure 5 shows that students taught by graduates of in-state programs that meet NCTQ's Comprehension standard make larger TRC gains than students taught by graduates of in-state programs that do

Figure 5: TRC Results for Ratings of Elementary Teacher Preparation Programs (Statewide Teacher Effectiveness Comparisons)

Note: *This figure displays results from models in which K-3 students' letter gain on the TRC is the outcome. The sample includes early-career teachers (<3 years of experience) in the 2014-15 through 2017-18 school years who are linked to an in-state elementary preparation program rated by NCTQ. Models control for student, classroom, teacher, and school covariates. '*' and '**' indicate statistical significance at the 0.05 and 0.01 levels, respectively.*

---

[7] TRC results for the UNC System vs. alternative entry teachers and for ECU and NCSU remain statistically significant in school fixed effect models; results for UNCCH and WSSU are no longer statistically significant when comparing within schools.

not meet the standard. This result is statistically significant across statewide and within-school comparisons but is not statistically significant for DIBELS. Conversely, students taught by graduates of in-state programs that meet NCTQ's Fluency standard have lower DIBELS scores and smaller gains on the TRC than students taught by graduates of in-state programs that do not meet the standard. These Fluency results are not statistically significant when comparing teachers within schools. Lastly, when comparing across all schools, we find that NCTQ's Early Literacy rating is not associated with the effectiveness of program graduates (DIBELS or TRC). However, when comparing within schools, we find a modest, positive relationship between Early Literacy ratings and students' TRC gains.
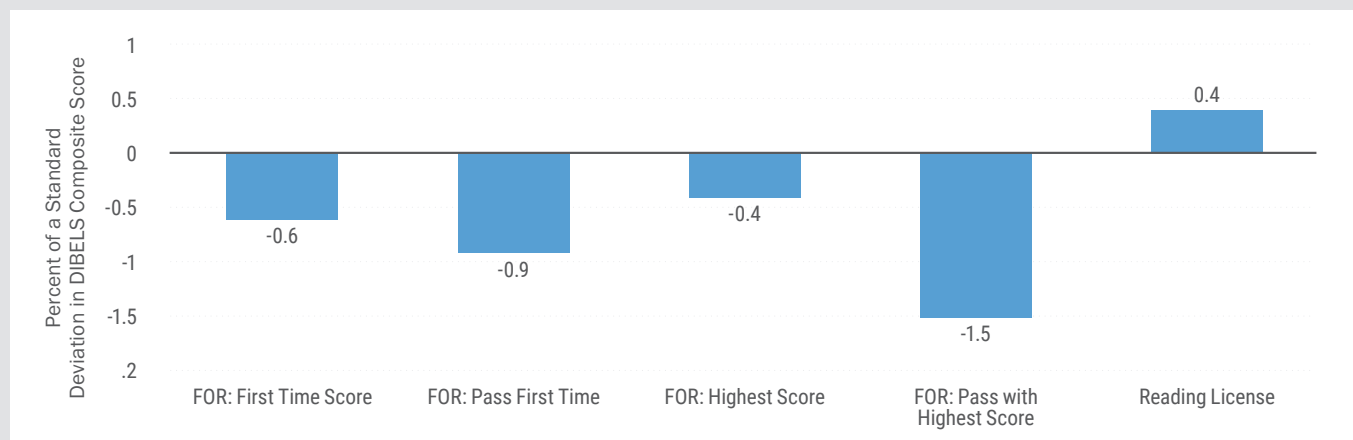
Do teacher exam scores and certification areas predict teacher effectiveness?

In these analyses, we focus on two measures: (1) teachers' scores from the Foundations of Reading licensure exam—both the total score and whether they passed the assessment[8]—and (2) an indicator for whether a teacher holds a reading license.[9] Each of these measures is a proxy

for teacher knowledge. The Foundations of Reading analyses indicate whether teachers who score higher on the exam, and presumably have greater literacy content knowledge, are more effective than those with lower scores. The reading license analyses indicate whether teachers who have taken a series of literacy courses, and presumably know more about reading instruction and assessment, are more effective than those who do not have a reading license.[10]

Figure 6 presents DIBELS results for the Foundations of Reading and reading licensure measures. Regarding the Foundations of Reading exam, we find no relationship between early-career teachers' exam scores and their students' DIBELS scores. Teachers who score higher or pass the exam are no more effective than peers who score lower or do not pass the exam. Likewise, we find that students taught by an early-career teacher with a reading license score no higher on DIBELS than comparable students taught by a teacher without a reading license. The one exception to this is low-performing students. Low-performing students taught by an early-career teacher with a reading license have higher end-of-year DIBELS composite scores than their low-performing peers taught by a teacher without a reading license.

Figure 6: DIBELS Results for Exam Scores and Holding a Reading License (Statewide Teacher Effectiveness Comparisons)
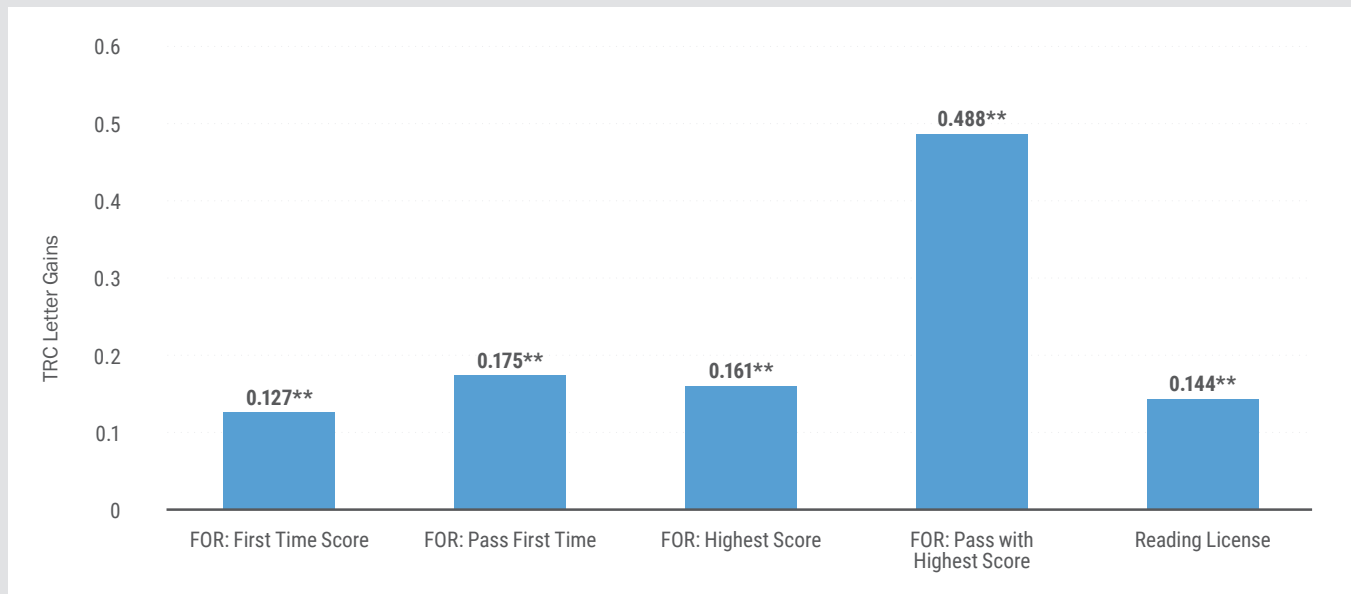

Note: *This figure displays results from models in which K-3 students' end-of-year DIBELS composite score (standardized within grade and year) is the outcome. The sample includes early-career teachers (<3 years of experience) in the 2014-15 through 2017-18 school years. Models control for student, classroom, teacher, and school covariates. '*' and '**' indicate statistical significance at the 0.05 and 0.01 levels, respectively. FOR=Foundations of Reading.*

[8] In our Foundations of Reading analyses, we assess data (the total score and an indicator for passing) from a teacher's first time taking the exam and from a teacher's highest score on the exam. For test-takers who pass on the first attempt these measures are the same.

[9] Approximately six percent of the early-career reading teachers in grades K–3 hold a reading license.

[10] Teachers earn a reading license in North Carolina by completing an undergraduate (add-on) program or graduate degree. These programs typically require 15–18 or 30–36 credit hours, respectively.

Figure 7: TRC Results for Exam Scores and Holding a Reading License (Statewide Teacher Effectiveness Comparisons)

Note: *This figure displays results from models in which K-3 students' letter gain on the TRC is the outcome. The sample includes early-career teachers (<3 years of experience) in the 2014-15 through 2017-18 school years. Models control for student, classroom, teacher, and school covariates. '*' and '**' indicate statistical significance at the 0.05 and 0.01 levels, respectively. FOR=Foundations of Reading.*

Figure 7 presents TRC results for the Foundations of Reading and reading licensure measures. Unlike the DIBELS results, we find strong relationships between our exam and licensure measures and students' TRC gains.[11] For example, students taught by an early-career teacher with Foundations of Reading scores one standard deviation higher gain 0.127 more reading levels. Likewise, students taught by early-career teachers who passed the Foundations of Reading exam on their first attempt gain 0.175 more reading levels than comparable students taught by a teacher who did not pass on the first attempt. Lastly, students taught by an early-career teacher with a reading license gain nearly 0.150 more reading levels on the TRC than comparable students taught by a teacher without a reading license. All Foundations of Reading and reading license results remain statistically significant in within school models.

Discussion

As North Carolina works to improve students' reading proficiency, *consistent* statewide achievement data is a valuable resource for informed decision making. We leveraged this statewide data to better understand the magnitude and correlates of teacher effectiveness in early grades reading. Our analyses return several key takeaways.

First, we show that teacher effects in early grades reading are large. Students taught by a highly effective teacher make much greater progress in acquiring essential fluency and comprehension skills. Put more powerfully, by the end of first grade, economically disadvantaged students are nearly two TRC reading levels behind their non-economically disadvantaged peers. Two additional reading levels is the difference between a highly effective and an ineffective teacher on the TRC. These findings should

---

[11] In previous analyses, EPIC found that Foundations of Reading scores predicted higher EVAAS estimates for teachers in K–2 reading (these EVAAS estimates come from TRC data). Foundations of Reading scores did not predict EVAAS estimates for the end-of-grade reading exam in grades 3-5.

be a catalyst for identifying highly effective teachers, understanding what practices and supports enable their success, and ensuring that more teachers develop the knowledge and skills necessary to succeed. To do this work, North Carolina needs valid and reliable K–3 reading assessments that allow stakeholders to quantify teachers' contributions to student learning.

Second, regarding educator preparation, our analyses identify areas for closer inspection and highlight the limitations of current knowledge. Several UNC System institutions had positive DIBELS and/or TRC results. These programs warrant further examination of their literacy practices. This investigation may be particularly important in the area of reading comprehension, where programs that met NCTQ's Comprehension standard had more effective graduates. More broadly, few of our educator preparation correlates predict teacher effectiveness—for the TRC and especially for DIBELS.

This finding suggests a need to better understand effective literacy preparation and to possess more granular measures of high-quality preparation practices.

Finally, we returned mixed findings for our Foundations of Reading and reading licensure measures. We find that these proxies for teacher knowledge predict teacher effectiveness on the TRC. These results are particularly important for state policy, as teachers must pass the Foundations of Reading exam to earn a continuing license. Conversely, neither of these measures predict teacher effectiveness on DIBELS. There are several possible explanations for this difference in results, including the content/skills assessed on DIBELS versus the TRC and the amount of subjectivity in TRC scoring. Understanding these differences across assessments may be critical to understanding teacher knowledge and making evidence-based policy decisions. Without such an understanding, these results should be interpreted cautiously.

For More Research on this Topic

Hernandez, D. (2011). Double Jeopardy: How Third-Grade Reading Skills and Poverty Influence High School Graduation. The Annie E. Casey Foundation. Available from: https://files.eric.ed.gov/fulltext/ED518818.pdf

Konstantopoulos, S. (2011). Teacher Effects in Early Grades: Evidence from a Randomized Study. *Teachers College Record, 113*(7), 1541-1565.

National Early Literacy Panel (2008). *Developing early literacy: Report of the National Early Literacy Panel.* National Institute for Literacy. https://lincs.ed.gov/publications/pdf/NELPReport09.pdf

National Institute of Child Health and Human Development, NIH, DHHS. (2000). *Report of the National Reading Panel: Teaching Children to Read* (00-4769). U.S. Government Printing Office. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf

Palardy, G.J. & Rumberger, R.W. (2008). Teacher Effectiveness in First Grade: The Importance of Background Qualifications, Attitudes, and Instructional Practices for Student Learning. *Educational Evaluation and Policy Analysis, 30*(2), 111-140.

Study Authors:

Kevin C. Bastian, Aubrey N. Comperatore, Paola Pilonieta, and Colleen E. Whittingham (July 2020)

UNC COLLEGE OF ARTS & SCIENCES PUBLIC POLICY

UNC CHARLOTTE